Creative Mathematics

The amazing workings of the mind

Everything is mathematical

Creative Mathematics

Creative Mathematics

The amazing workings of the mind

Miquel Albertí

Everything is mathematical

Published by RBA Coleccionables, S.A.
c/o Hothouse Developments Ltd
91 Brick Lane, London, E1 6QL

Localisation: Windmill Books Ltd.

ISSN: 2050-649X

Printed in Spain

Contents

Preface

With the same rules that constrain a beginner, a seasoned chess player develops strategies that at first seem incredible. But anyone can learn to play chess. And the only way to learn is by playing, because playing is much more than moving the pieces according to the rules. When you play, you create.

Thousands of years ago someone had the idea of making some marks on a bone or a stone, each of which corresponded to a concept in their mind. The outline or shape of those marks was irrelevant; what was important was their relationships. A specific thought in my mind shows up as a mark seemingly identical to the others. Later, names were given to each symbol and to each group of symbols. This allowed different groupings to be distinguished and which one was greater than another to be agreed. The number is, without doubt, the greatest creation of mathematics and, together with the word, the greatest human creation.

The other great mathematical creation is its production system. Each result is validated by an expert community which examines it in detail. If it has imperfections, they are corrected. That correction is governed through the logic of human understanding. At the end a proven result is obtained, that is, a result which is deducible by anyone who wishes to go through the series of links elaborated by the author. The result is called a theorem.

Traditionally, mathematicians have maintained a tacit pact not to show their errors, nonsense or failed insights. Mathematical products are published unblemished. This also forms part of other traditions. Craftsmen only display their work when it is finished, but everyone knows that in order to achieve such beauty, many hours of work are required. That work gives value to the product. The carving was not created overnight; it was a series of tests and mistakes, corrections and restarts.

Sometimes we get the feeling that in mathematics the theorems are obtained from other previous theorems. The merit is in having the intellectual capacity to be able to create good combinations and apply the rules of logic well. But logic on its own is barren, it does not produce anything without an agent to activate it, and the activation of mathematical development is all about insight, analogies, tests and debates. This is where mathematical creativity comes in.

To create means to produce something new and unknown; that is why creation is closely linked to learning. Starting with the premise that to know mathematics is to know how to use it, mathematical creation is based on proposing good questions and resolving problems. That makes a professional mathematician. Each demonstrated

theorem is not the end of the process, but a link that generates new conundrums or which allows other problems and conjectures to be solved. Creativity is also found in the ability to pose new questions without answers.

The version of mathematical creation proposed here is not limited just to the professionals. Anyone can create mathematics. It is possible that what is created by the amateur is not new to the professional, but it is to its originator. Perhaps it is not inspired by theorems and problems of the professional, but it may be inspired by the amateur's everyday life, in what he or she sees, in his or her work – and play. For this to occur we must look at both mathematics and reality in a different way.

Surprisingly, mathematical creations do not always end in satisfaction. Throughout history there have been traumatic creations which have been triggered by great crises. If we believe that numbers serve to count units and that the relationship between everything in the universe can be established by means of proportions between simple numbers, what is the square root of two? What is a negative number? And the square root of minus one? Creativity produces monsters which have to be managed and digested through changes in approach. After all, we do not look at a Picasso in the same way we look at a Velázquez. We do not listen to Stravinski or Miles Davis in the same way we listen to Bach or Händel.

How do we create? Are there guidelines for the creative process? The most common belief is that a mathematical creator has happy accidents, moments of extraordinary inspiration that allow him or her to find the key to a problem. It can be said that a genuine mathematician has a gift that others lack and which is activated in the face of difficulty. There is a 'click' in his or her mind and a light turns on. As in any other discipline, some people are better than others at mathematics. However, our intent here is to emphasize the characteristics and guidelines of creation and to see that many of them are achievable for all.

We will start by looking at how some of history's greatest mathematical creations have been related to times of crisis. Then we will try to clear up the notion of serendipity in the solving of problems, and we will see that it has strong ties to the learning process. Later there will be a few examples of mathematical creativity derived from activities from anyone's everyday life. We will dedicate an entire chapter to this experience aspect of the mathematical routine, which will include the history of the development of mathematical knowledge by the author based on an intercultural and extra-academic experience. This part also illustrates one of the book's most important theories: culture and society play a fundamental role in mathematical creativity and in the mathematics that is created.

In the penultimate chapter of the book we will switch the focus and go from talking about creativity in mathematics to the mathematics of creativity, as understood by professionals in traditionally creative fields, such as the world of design and publicity. And we will conclude with a recap all the previous chapters in search of a unique characterisation of mathematical creation in order to identify the fundamental guidelines for discovering and nurturing it.

Chapter 1

The Pillars of Mathematical Creation

Mathematics is commonly characterised as an 'exact science'. The emphasis on the adjective was so common that the noun to which it referred was overlooked. When someone studied mathematics at university, they simply used to study 'exactness'.

This has been the paradigm of mathematics: exactness, precision, rigour, the banishment of error and indecision, the absence of middle-ground, black and white without nuances, straight or curved lines, finite or infinite, open or closed, correct or incorrect, good or bad. This is an itinerary guided, with irrefutable authority, by logic, and applied in turn to a few principles that are as rudimentary and universal – or, at least, they appear to be – as life itself.

This model is based on an ancient work which would become the book of mathematics par excellence, both in terms of its content and its form of communication: Euclid's *Elements*. From essential truths taken as true (postulates) other new ones are deduced (theorems), less evident than the first, and which lead in turn to others which are even less certain. The chain of results obtained derives from a body of fundamentally geometric knowledge, the truth of which is guaranteed by logical deduction. No result is obtained on a whim but as the consequence of logical reasoning based on the primeval postulates.

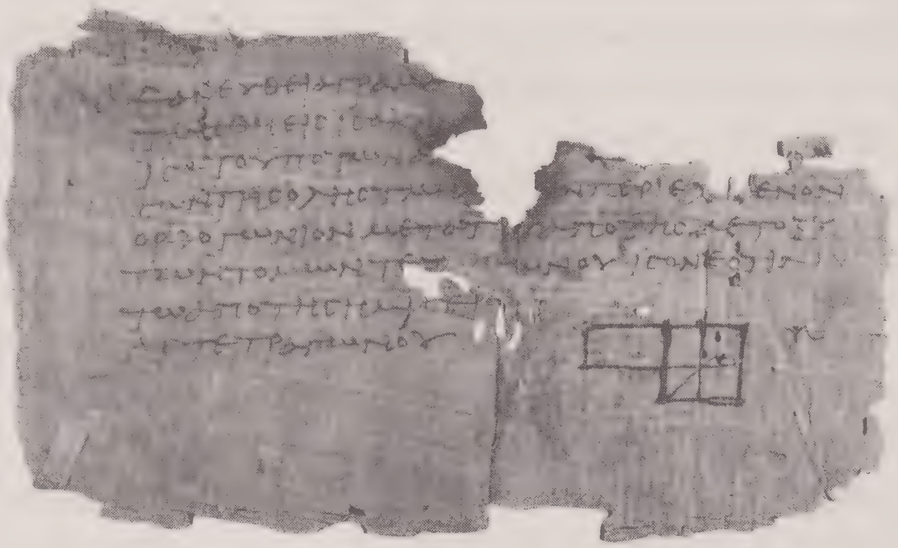
Until recently, Euclid's work had also been the model followed by mathematical education. From there, the most common conception of mathematics is linked to this ideal relationship between exact results bound to one another as per the everlasting pattern of axiom-theorem-demonstration-corollary-exercise. That was mathematics, it was taught like that, it was learnt like that, it was perceived like that.

However, as intellectually rigorous and intelligently skilled as Euclid may have been, can anyone imagine that the *Elements* were obtained on an impulse once the pillars of the postulates had been established? Could there not have been errors in the connections? Why illustrate the arguments with diagrams? Do they not include suppositions that are not included in the postulates? The answer is of course yes. In the first theorem, Euclid, assumes that two arcs of a circle intersect each other at a

EUCLID, THE MAN OF THE METHOD

Little is known about the creator of the greatest paradigms, both of books and methods, in mathematics. We know that he lived in around 300 BC and that he taught in Alexandria, Egypt. His best-known work is, of course, *Elements*, made up of three books with more than four hundred proposals deduced from five postulates, five common notions or axioms, and 132 definitions. Examples of these different types of statement are:

Def. 1: A point is that which has no parts.
Def. 2: A line is a length without width.
Def. 3: The ends of a line are points.
Pos. 1: A straight line segment can be drawn joining any two points.
Pos. 2: And any straight line segment can be extended indefinitely, to form a straight line.
Pos. 3: And given any straight line segment, a circle can be drawn having the segment as radius and one endpoint as centre.
Axiom 1: Things that equal the same thing also equal one another.
Axiom 2: If equals are added to equals, then the wholes are equal.



Papyrus with a fragment of Proposal No. 5 from 'Book II'.

point. But which postulate guarantees that fact? Are they continuous lines? If we amplify the line drawn with a pencil under a microscope, will empty spaces not appear? Also, what is Euclid thinking about, segments of lines or indefinite ones?

Segments, straight lines, triangles, squares, circles, etc.... Mathematics addresses figures of such perfection that they cannot be of human creation, only divine. Or, excluding the latter possibility, they exist for themselves in an ideal world of immaculate perfection. This perspective, clearly a Platonic aftertaste, has captivated practically all mathematicians since ancient times. Those who have made contributions on this basis have done so after years of dedication and reflection on the essence of their activity. For the Pythagoreans, numeric proportion governed the laws of everything in existence. If perfection was close to numbers, then numbers were the closest thing to God. The geometric circle is a perfect entity with properties that are said to have been 'discovered'. In some way this geometric figure exists in everybody's shared imagination, and by studying it we are simply discovering its properties and its relationships with other ideal figures. Here we have the traditional idea of mathematical discovery. We will question it later.

Logic does not create, but we cannot create without it

Logic is vital in mathematics. Justification of mathematical results is based on it, it is the determinant judge of their truth or falsehood. But mathematics are not limited to it; not all mathematics is logical. If the theorems were obtained by simply applying the formal rules, we could give a computer the job of endlessly producing them. The problem is that what is normally revealed of the mathematical world is already complete, tidy results.

A long time had to pass until the world of maths was honest with itself and recognised that, behind closed doors, mathematics is something else. Mathematical dishes continue to be served on clean plates, balanced and without the slightest defect. Many chefs from the community of scholars try them over and over again before giving the go-ahead to make them public. They search for errors and, if they find any, they correct them. Any dish with an unrecoverable error is immediately rejected and returned to the kitchen. Therefore, the daily mathematical work does not take place in the restaurant's dining room, but out of sight in the kitchen. This is where the axioms, theories, demonstrations, etc, are being cooked up. This is where errors are made, tests are carried out and ideas and demonstrations are disproved. The chefs get covered in grease and stains and they despair when logic refuses to give in to the desires of their intuition. So they continuously curse the activity which many people assume to be handed down from an ethereal, perhaps divine, realm beyond reason.

But it is not the fire of logic, or not only the fire of logic, which feeds the mathematical kitchen. It is also intuition, analogy and experimentation, conjecture, in other words, human thought. And given that not everyone thinks the same or has the same interests, the society and culture of mathematics also influences its activities and resources. What makes one theorem more valuable than another? Why try to prove some and ignore others? Logic produces an infinity of unimportant trivialities. The development of mathematical knowledge is a result of the human being's interest in solving problems, be they theoretical or practical, useful or useless, both inspired by the thirst for knowledge and the personal challenge of creativity.

This more complex and precise picture of the nature of both mathematics and mathematical activity is demonstrated in a series of ground-breaking books, such as *What is Mathematics?* (1941), by Americans Richard Courant and Herbert Robbins; and the subsequent *The Mathematical Experience* (1999) by their compatriots Philip J. Davis and Reuben Hersh, or in the latter's book *What is Mathematics, Really?* (1997). In it, Hersh makes a simple and clear observation: "The formula 2+2=4 can be proven as a theorem in a formal axiomatic system, but it derives its force and conviction from its physical model of collecting coins or pebbles." Moreover, the logic that allows this formal demonstration to which Hersh is referring came long after the collection of pebbles. Courant and Robbins, for their part, insist on the important role that experimentation, intuition and analogy played in the development of mathematical knowledge:

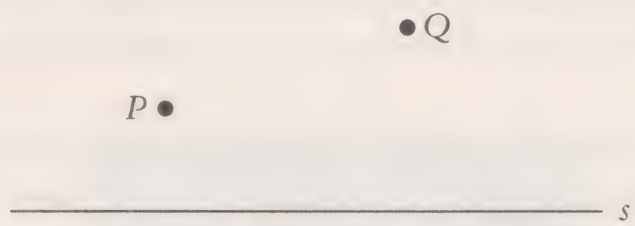
> "...Although the principle of mathematical induction suffices to prove the formula, once this formula has been written down, the proof gives no indication of how this formula was arrived at... The fact that the proof of a theorem consists in the application of certain simple rules of logic does not dispose of the creative element in mathematics, which lies in the choice of the possibilities to be examined. The question of the origin of the hypothesis belongs to a domain in which no particularly general rules can be given; experiment, analogy, and constructive intuition play their part here."

In mathematics logic plays a fundamental role, but with a weaker relationship with discovery or invention than is apparent. Logic does not choose paths or determine the way to seek a solution. Experimentation, analogy and intuition open paths which logic will later pave and convert into roads open to anything that wants to travel that way.

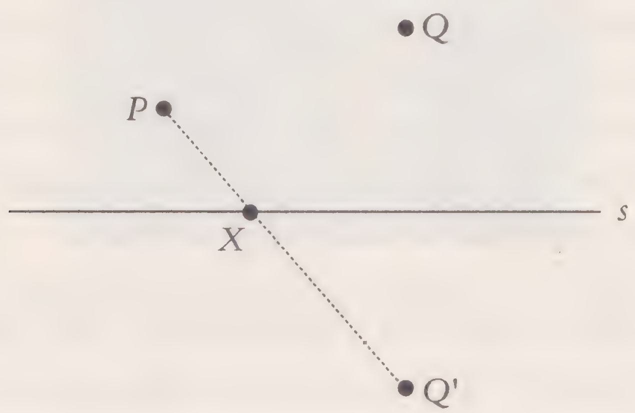
A good example of this is given below; a celebrated geometric problem solved elegantly with serendipity, or what we will call a 'happy idea'.

The happy idea

We have two points, P and Q, and a segment s, as shown in the diagram. We want to get from P to Q passing through a point on s. Which point on s gives the shortest route?

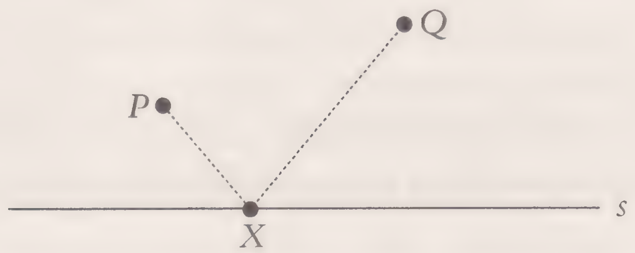


In order to resolve this problem let's imagine segment s as if it were a mirror. Let's draw the reflection of Q with respect to it and call it Q'. Then draw the segment that joins P to Q' and intersects with s at point X:



The segment PQ' determines the shortest route between P and Q', and its intersection with s, point X, is the point that we would have to pass through in order to travel along it. All that remains now is to use the symmetry again, reflecting segment XQ' in mirror s and seeing that line XQ has the same length as XQ'.

Thus the solution is obtained, the broken line PXQ, of length identical to that of line PQ':



Therefore, the shortest path for going from P to Q passing through s is to move towards point X.

Is the above idea, based on symmetry, a happy idea? Any effective idea that has not occurred to us could merit that description. However, and this is one of the central points of this book: the generation of happy ideas – mathematical creativity itself – can be taught.

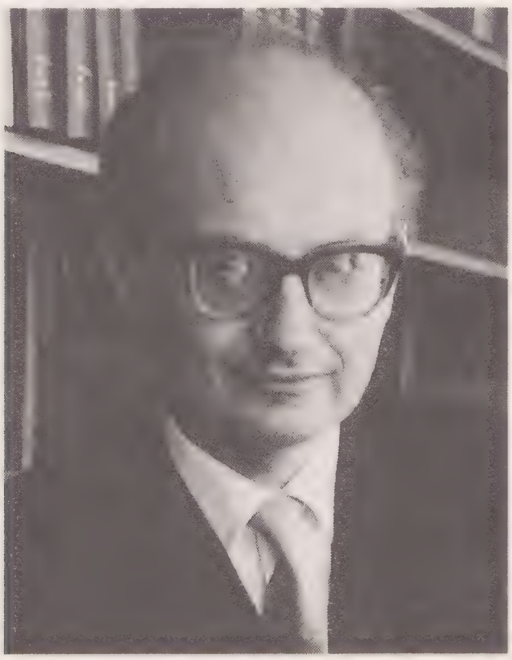
A well-argued and logical demonstration of this solution is based on the fact that symmetry conserves distances and that a segment is the shortest line between two points on a plane. The solution may even appear simple and trivial once shown, but it is difficult to imagine for someone who is seeing the problem for the first time. This is an example of creativity. Logic, on its own, does not lead to the result. It is achieved through the insight to imagine additional lines which are not drawn on the given at the start and to use them to create relationships between the various elements of the diagram. Logic enables many actions, but it does not give us arguments for choosing the best approach.

This creative capacity in mathematics is not universal, just as artistic, musical, architectural and scientific creativity are not either. But it is that which many people have been calling the 'happy idea', a kind of trick or subtle, almost magical, inspiration that did not figure in the data or in the problem given and which is not immediately apparent in the initial proposal.

Happy ideas exist, but they are not restricted to the geniuses, and not all problems are resolved with them. As we will see later, they are normally the fruits of intense and continuous work, as well as the search for relationships between the elements of a problem. Of the practically countless relationships that can be expressed between the data or elements in a problem, how can we choose the one that resolves it? In the choice of these 'good possibilities' resides mathematical creativity.

Humanistic, social and cultural components of mathematics

The most humanistic vision of mathematics considers it to be a historical, social and cultural product. Indeed, many of the results prepared by the mathematician in his 'kitchen' are developed in the same way as scientific theories – by means of "proofs and refutations" in the words of Imre Lakatos. The mathematician clears a path in the jungle, overcomes obstacles and creates, step by step, counterexample by counterexample, his theorem. What happens is that mathematical theories are presented and justified by means of impeccable logical reasoning, putting together what to everyone's eyes looks like a smooth and safe highway to the truth.



Hungarian mathematician and philosopher of the origin of science, Imre Lakatos.

However, other hidden factors, such as experimentation, intuition and analogy are also involved in paving the way. Hersh again:

> "Real-life proof is informal, in whole or in part. A piece of formal argument – a calculation – is meaningful only to complete or verify some informal reasoning. The formal logic of proof is a topic for study in logic rather than a truthful picture of real-life mathematics..."

The mathematical community validates this as 'knowledge' by revising and criticising the results presented by its members. This phase resides in empirical practice and in perceptive experiences and is similar in every way to those of any other individual who relates to their surroundings. This 'naturalist' paradigm provides – as we will see throughout the book – the possibility of finding mathematical knowledge in activities that are not directly linked to the academic field.

The interpretation of the nature of mathematics as a cultural product, as reliable as any other, the basis of whose justification is empirical, is known as 'social constructivism'; it is a concept that would be familiar to the above-mentioned authors, such as Lakatos, Davis and Hersh. In the words of one of the best-known representatives of this school of thought, American Paul Ernest:

> "In summary, the social constructivist thesis is that objective knowledge of mathematics exists in and through the social world of human action, interaction and rules, supported by individuals' subjective knowledge of mathematics which needs constant re-creation. Thus, subjective knowledge recreates objective knowledge, without the latter reducible to the former."

This is a vision of mathematics where science and education go hand in hand, and in which the teaching of the discipline is determined by the subject and the society or culture to which it belongs. Traditionally, the only non-Western social and cultural environments mentioned by mathematics historians were always those of the ancient world: Mesopotamia, ancient Egypt, ancient Greece, the early Islamic era, ancient India, and the ancient Chinese world. All are now gone; we never consider living worlds, societies or cultures.

Paradoxically, historians agree that the existence of mathematical knowledge dates back to periods as remote as the origin of language, when what we would today consider civilisation did not exist, let alone Western culture. So, if we consider that the efforts of the human race to establish numbering systems constituted the beginnings of mathematical activity and, in agreement with the best-known science historians, that the origins of this activity are more ancient than writing, we should conclude that not only was mathematics created outside of our culture, it was also created long before its rise.

Is mathematics created or discovered?

The majesty of some mathematical results and Plato's inheritance have inspired the idea of discovery, but mathematics is not discovered, it is created. In the words of Spanish logistician Josep Pla:

> "...Mathematics, almost like language, is a *construct* of the human mind with its own vitality, which makes us think that it exists regardless of its knowledge and its creation. This, if you will excuse my emphasis, is erroneous."

This statement can be illustrated by means of an example. Let's suppose that various animals come to a pool to drink. If one person watches them, he could describe them in many different ways and formulate a lot of questions about them. However, all of them would be inspired by his culture. Some formalist mathematicians would argue that there are seven animals drinking and that there are seven of them regardless of the observer. The perspective championed here would argue that the seven is determined by cultural matters, because anyone who is observing knows how to count, because he knows how to discern between few and many, because he is interested in elucidating how few and how many there are. However, perhaps that person who counts seven animals is overlooking a fundamental fact, which is there before his eyes or his senses, but which he does not manage to see because his different culture prevents him for formulating other questions to those which he mutters in his language. How can we know if those other questions are also of a mathematical nature and as relevant or more so than the precise quantity of thirsty organisms.

Therefore, it is reasonable to state, as did Hersh and Ernest, that the mathematics that we know is a social and cultural product and, therefore, that different cultures can create different mathematics. In fact, this has already happened. Is non-Euclidean geometry, developed on the escritoires of bourgeois Europe of the 18th century, not different to that of Euclid's classical geometry traced in the earth of ancient Greece 2,300 years before?

All Euclidean mathematics has a finite character. In it there are neither iterative procedures nor any concept of limit. In a similar context, differential calculus would not have been considered to be mathematical. Today the globalisation of mathematics is such that it has engulfed practically all of the differences. Euclidean, projective, spherical and fractal geometry, resolutions with finite and recurrent methods, the use of basic technology (a ruler and compass) or advanced software, we refer to

all of this, and much more, with the same name: mathematics. Now it is just one thing, but previously it was not.

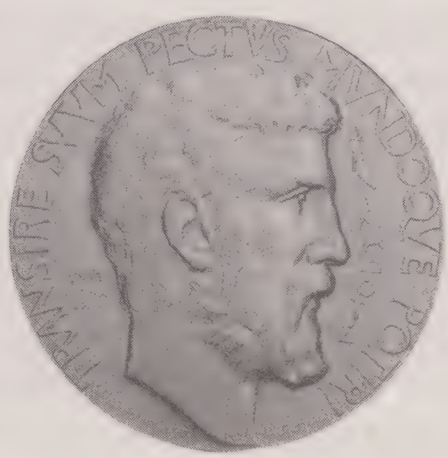
In the bath with Archimedes and Poincaré

Legend has it that it was mid-dip, during the everyday activity of having a bath, when the great scholar and mathematician Archimedes shouted "Eureka!" when he came up with the (happy?) idea that the volume of a submerged body (not necessarily a human one) is the equivalent to that of the water which it displaces. This idea of happy and spontaneous inspiration has been and continues to be the paradigm of mathematical creation. However, its spontaneity is only apparent. Great mathematicians such as Frenchman Henri Poincaré have also had similar experiences, and have spoken about the how and where of that creative and far-sighted inspiration.

How is an extraordinary idea summoned in a person's mind? What activities does it provoke? The answers to these questions should not be sought in mathematics, but in psychology. It was Poincaré who offered a description of how the mind of a mathematician works, and he did it in front of the Paris Psychology Society at the beginning of the last century.

ARCHIMEDES OF SYRACUSE (287–212 BC)

This genius of old died at the hands of a Roman soldier, despite general Marcelo's orders to keep him alive. According to legend, his executioner had no sympathy for someone who continued to be absorbed in mathematical tasks while his house was invaded by the troops who had besieged the city. His work includes: the law of the lever, the hydrostatic principle, the approximation of the area of a circle, the trisection of an angle, the area of parabolic and spherical segments, and the book on the sphere and the cylinder.



The bust of Archimedes figures on the Fields Medal, which is awarded every four years to the best mathematicians under 40 years old. It is considered the Nobel Prize of mathematics.

Poincaré started his presentation by proposing two paradoxical questions: "How can there be anyone who does not understand or resists mathematics? How is error possible in mathematics?"

The first query arises when pointing out that mathematics is based on a logic that is accepted by the whole world and on principles that are so basic and common that only a madman would reject them. The second appears when one thinks that a mathematician is someone intelligent who works following the rules of logic and that, therefore, they should not make mistakes. Applying the rules of the game you end up winning. Also, there are people who are very capable when it comes to the logic of ordinary acts, but are incapable of following the thread of a mathematical demonstration made up of small, even briefer and utterly logical steps.

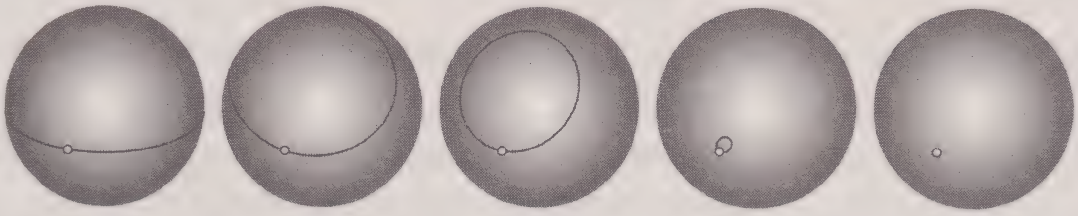
A crucial observation is that a mathematical demonstration is not a juxtaposition of syllogisms placed in a certain order, and the order in which they are placed is much more important than the elements themselves. If one has the intuition of that order, they no longer have to fear forgetting any of its elements, since for them, memory and concentration are enough. But that feeling or intuition for perceiving hidden or shared relationships or harmonies between apparently different things does not appear to be a very common thing in the world. And, according to Poincaré, this is the ability to tell apart those who can create mathematics and those who can learn it, understand it and apply it.

Mathematical creation does not consist of combining previous knowledge. A computer could do that, but many of those combinations would not be interesting. To Poincaré to create is to choose the useful combinations, which are very few, from among the useless ones, which are overly abundant.

Poincaré's creative process is summarised in phases. It starts with arduous and prolonged work on the subject (a couple of weeks). Then an unfamiliar event (having a strong coffee late in the day) prevents him from sleeping, and ideas hit him during his state of wakefulness. Then two ideas are linked and they are established. Later the results are written down. This is followed by a conscious and deliberate idea guided by an analogy. He enters into another phase in which he carries out an activity which has nothing to do with mathematics (a rock survey) and in which he can forget about the mathematical work. Right in the middle of carrying out an every-day activity (getting onto a bus) he comes up with a key relationship between two apparently unrelated elements (Fuchsian functions and non-Euclidean geometry). Back at home, he checks that the result is satisfactory.

HENRI POINCARÉ (1854-1912)

This French mathematician is mostly celebrated for the topological conjecture that carries his name. This was eventually demonstrated, at least in its salient points, by Russian mathematician Grigori Perelman in 2002. A loop can be continuously tightened to a point on the bi-dimensional surface of a sphere. The Poincaré conjecture states that there is a similar situation to characterise a three-dimensional surface in four-dimensional space.



The illustration shows a loop tightening around a point on the sphere.

Poincaré's sudden illumination corresponds to a long period of intense, conscious and unconscious mental activity. And this unconscious work, sometimes more productive than the conscious work, only seems to be activated after a period of intense work, as if we had left the computer in *sleep* mode or minimised a window on the computer to open another program and work in other applications. But the program on standby, the minimised app, continues working and comes up with a solution of which we will only be aware when we reactivate it, by displaying it again, be it with a voluntary click or simply closing all the other programs and windows. Poincaré emphasised the role of making strenuous voluntary efforts, even though they were fruitless. Without them nothing else would happen.

We do not know which mental processes led to Archimedes' great results, but it was probably something similar to that undertaken by Poincaré. Anyone who has worked in mathematics, as professionals or amateurs, will also be able to refer to similar experiences.

This is the traditional psychological perspective on mathematical creation. But there are a few other noteworthy things in the creative experiences of the great French mathematician. One is the fact that his farsightedness was centred around linking apparently different things. This aspect was fundamental in his life, to the point of stating that mathematics consists of giving the same name to different things. Such capacity is not only well-known in mathematical creation, but also characteristic to all things creative. The other fact, which the previous one is based on and which Poincaré, Courant and Robbins (1996), Pólya (1988) and Lakatos (1994) all focus on, is the role played by analogy in mathematical creation.

We have thus identified a determinant aspect in mathematical creation: analogy. At this point we could say: "Do you want to create mathematics? Then think of analogies and put away logic." Which other activities affect creativity?

Psychology of creativity

The psychological perspective on thought also echoes the divergence between logical thought and creative thought, as it confirms the existence of a mental activity differs from the ability to deduce conclusions based on well-defined principles and rules. It is altogether different, however, to specify what creation will lead to valuable innovations. Socrates, through Plato, proposed this paradox:

> "How are you going to look for something when you don't in the least know what it is? Which of the things you do not know is the one which you propose to look for? And if you chance upon it, how will you know it is what you are looking for, given that you do not know what it is?"

We should not disregard the possibility that randomness may end the search, with the arrival of a link that resolves a problem, the good combination from among all the countless ones to which Poincaré referred. However, something tells us that chance is not responsible for the idea. At least, not only chance.

Psychologists draw a distinction between four phases of the creative process:

1. Preparation.
2. Incubation.
3. Illumination.
4. Verification.

In the first we get acquainted with the problem, specifying the question, gathering data, writing the formulation and sizing up possible solutions, strategies, relationships, etc. The second phase, incubation, is subconscious. In it, unexpected associations are produced that may provide us with the best-trodden and common routes. From here illumination emerges, which appears to sprout spontaneously like divine inspiration. A light switches on, illuminating the 'happy idea' which resolves the problem. A revelation is produced. Ecstasy may result as was the case with Charles Darwin, who was struck with his theory of natural selection during a drive in a carriage.

Creativity is the equivalent to mental fluidity, and there are a lot of psychologists who highlight the role of associations or combinations of ideas in the course of that fluidity. However, creative essence does not reside in the production of associations, but in the "criterion for differentiating the trivial ones from the genuinely good ones" (Pinillos, 1981: 469). Psychologists are in agreement with Poincaré. According to them, creative activity consists of a special form of solving problems characterised by its newness, lack of conventionalism, persistence and effort of the resulting process. There are also people who believe that this newness should be unusual, with little relation to anything before it and its acute contributions.

The characteristics of creativity are so varied that they would be said to be boundless. Originality and innovation are subject to historical criteria, they are never absolute and their appreciation changes with time and culture. Their marked subjectivity impedes the establishment of a precise criterion for originality which is in detriment to their investigation. "Logic, experience and experimental analysis are, of course, essential elements of the creative thinker. But creativity is something else" (Matussek, 1977). Here we find new elements related to creation: logic, experimentation and practice.

The creator does not stop thinking, while the non-creator clings to what he has just thought and feels satisfied that he does not have to continue thinking. He finds it difficult to move from one representation to the next. In a creator, ideas move easily from one field to another and they can adopt various approaches at the same time without clinging to any of them. The ability to create new definitions is fundamental for the understanding of things, as we can only talk about true knowledge when an idea has been understood properly. Many things glimpsed, lived and experienced continue to be unknown because they have still not been understood. Creators can 'problemise' things and form causal links with greater ease than non-creators. They present them as a problem and thus look for solutions.

We can see that a new factor has appeared, comprehension. In the case of mathematics this can be paramount. Anyone who does not understand a problem cannot resolve it, and perhaps the sudden inspiration that we spoke of before corresponds to that moment of clarity through which we end up understanding the phenomenon we are studying. Thus, cries of "Eureka!" should be replaced with another, more profound one: "I understand!" The bases for the field in which analogy, experimentation, practice, logic, comprehension and the approach to problems are found have thus been established. These are fundamental components of heuristics.

Heuristics: the melting pot of mathematical creation

With practice, the ability to see suitable relationships can be developed. The first thing to do is to try various alternatives, test and make mistakes, start again and try other ways. In other words, experiment. This is how we learn to select useful paths and reject the useless ones without the need to exhaustively go down each one. This art of invention, of creating meaningful paths to the solution of a mathematical problem is known by the name of 'heuristics'.

Its main champion, and the person who rescued it from oblivion as a technique for mathematical development, was the Hungarian George Pólya, working in the first half of the 20th century. He said: "Mathematics has two faces. It is the rigorous science of Euclid, but it is also something else. Mathematics presented in the Euclidean way appears as a systematic, deductive science; but mathematics in the making appears as an experimental, inductive science. Both aspects are as old as the science of mathematics itself." It is precisely this 'something else' which has a lot to do with mathematical creativity, as will be made apparent later. In *How to Solve It*, Pólya provides four fundamental steps for solving a mathematical problem:

1. Understand the problem.
2. Create a plan to solve it.
3. Carry out the plan.
4. Examine the obtained solution and revise the process.

Pólya distinguishes between problems to be demonstrated and problems to be found. The problem dealt with in the previous section was of the second type. At the end of this chapter is the heuristic solution to a problem of the first type.

GEORGE PÓLYA (1887–1985)

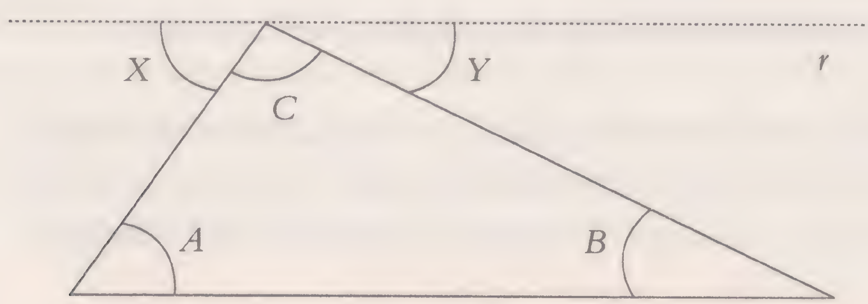
This Hungarian mathematician developed fundamental techniques for the resolution of problems. The Pólya conjecture, proposed in 1919, states that most natural numbers less than any given number have an odd number of prime factors. This conjecture was proved false in 1950, although the smallest counterexample was found in 1980; it is the number 906,150,257.

The creative character of the heuristic method has been highlighted by Davis and Hersh: "The heuristic example of demonstrations and refutations proposed by Lakatos...can be applied by individuals in their efforts to create new mathematics." What happens is that in order to do just that, a change of focus and a good dose of courage is necessary, given that the long-established idea of what mathematics is and how it works differs greatly from its actual reality.

Let's take a look at an example of one of the fundamental theories of plane geometry: the sum of the three angles of a triangle gives a straight angle. That is, in any triangle with angles A, B and C it is true that $A+B+C=180°$, 180° being the size of a straight angle.

To this end, the three angles would have to be brought together on one vertex to see that their union produces a straight angle. This fact can be proved experimentally cutting the three corners of a paper triangle and reorganising the three pieces at a single point. But this does not prove the phenomenon. It just shows that it is valid in this particular case.

The best-known demonstration of this phenomenon is based on drawing a straight line r on the vertex of a triangle, for example C, parallel to its opposite side. Thus, two angles exterior to the triangle are created on vertex C. Let's call them X and Y:

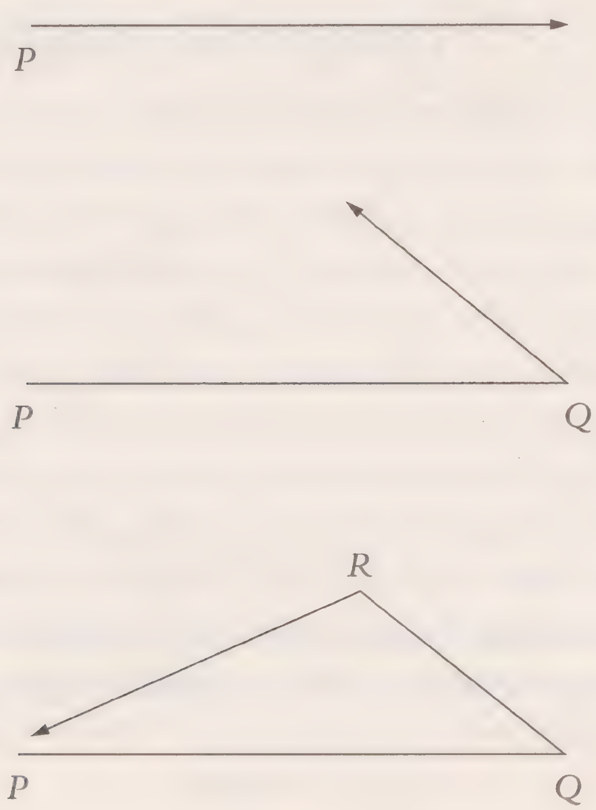


But given r is parallel with side AB, angle X is equal to angle A. The same applies to B and Y. But, evidently, the sum of the three angles on vertex C equals the straight angle of straight line r. Therefore, $180° = X + C + Y = A + C + B$, and the sum of the three angles of the triangle equals a straight angle: $A + B + C = 180°$.

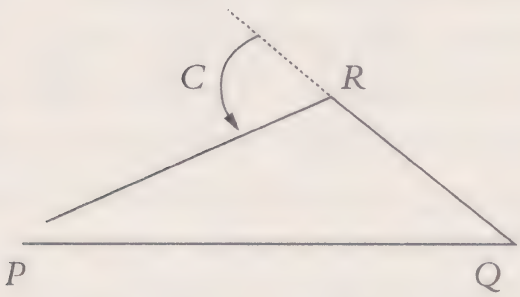
This demonstration is based on drawing an additional line to support the argument. We could think of many things and add points and lines to the diagram, but not all of them would be as useful. The shape and position of the triangle being drawn, that is, the situation we see before our eyes, also influence the resolution of the problem.

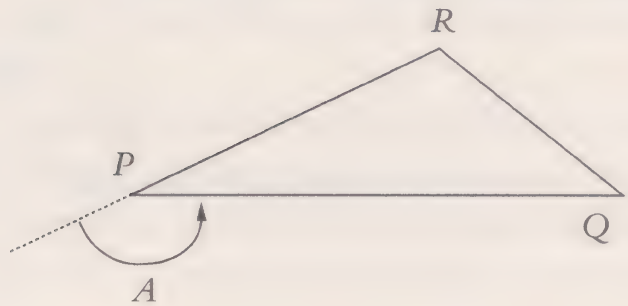
Anyone to whom this happy idea of drawing a line parallel to another vertex does not occur can try and focus on the demonstration in another way. For this, let's reflect a little on what we are doing by drawing a triangle.

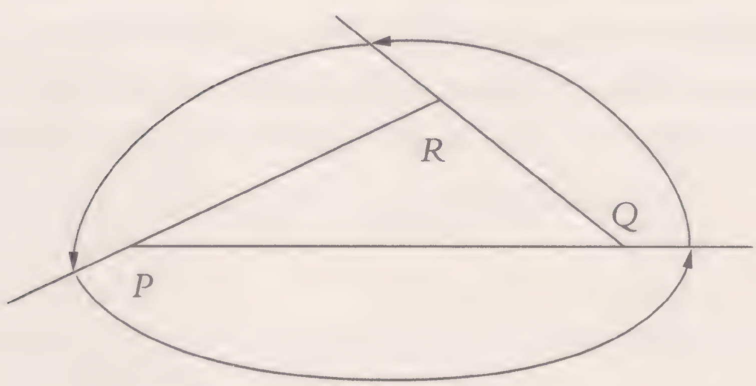
Drawing a triangle means drawing a closed line with three corners. To do so we put the tip of the pencil on a point P of the paper and we glide it in a straight line, let's say, to the right. End Q of that segment is determined by a change in direction of the line which will be repeated again at the third vertex R. From there we turn again in order to join up with start point P:



Now let's establish the roots of our demonstration. The triangle is composed of the outline of three segments. The second of which starts at Q with a rotation of angle B. Then we continue until R where we rotate again at an angle of C with which we meet starting point P. Finally, upon arriving at P we rotate by angle A in order to situate ourselves in the same direction in which we started the drawing:







As we have returned to the same direction (that of segment PQ) with which we started, the three changes of direction total a complete rotation of 360°:

$$A+B+C=360°.$$

But each one of the three angles is supplementary to the interior angle of each vertex, in other words, what is missing from A, B and C to make 180°. Calling those interior angles α, β and γ we get:

$$(180^\circ - \alpha) + (180^\circ - \beta) + (180^\circ - \gamma) = A + B + C = 360^\circ.$$
$$540^\circ - (\alpha + \beta + \gamma) = 360^\circ.$$
$$180^\circ = \alpha + \beta + \gamma.$$

And the three angles of the triangle total a straight angle.

This demonstration could be called 'situated', as it is based directly on the way in which the figure is drawn, that is, the situation. Nothing additional is required other than the reflection on the action being carried out.

Pólya and Lakatos' heuristic method is the way in which most mathematical knowledge has actually been developed throughout history yet has nearly always remained hidden. The demonstration of a theory is managed by first local, then global criticisms and counterexamples, which not only have the purpose of refining it, but also to rewrite the formulation in order to create new definitions and categories. The object of the process consists in reaching a more precise or general formulation and a definitive demonstration which allows the comprehension of the theorem.

Mathematical education and creativity

Constructivist mathematical education attempts to show that with something we already know and with some help and good management of the information available we are capable of solving problems. Meanwhile, we learn by creating mathematics that will be our mathematics. Perhaps it is not new knowledge on a world level, but we will have carried out a process that is practically identical to that of the professional mathematician. We should undoubtedly feel as delighted as he or she. The experience of teachers guiding students, shows that this is very much the case.

So, if everyone can create something from mathematics, what's the problem? It's our commitment to the cause. Learning is looking for answers but then we need questions. Who creates mathematical questions about what they see, hear, do or experience? Practically nobody. That is what sets mathematicians apart from non-mathematicians. However, the world cannot be understood without mathematics. Looking at something mathematically, proposing problems and resolving them is also a learnt skill.

It is impossible – so far – to determine what the brain is doing in the incubation phase to resolve a problem or create something new. We cannot control that moment of revelation either. By definition it is produced spontaneously, without forewarning and unconsciously. What we can influence is the other two phases.

There is not much to say regarding the period of verification. It involves the very traditional activities of mathematics: routine checking of result, ensuring they answer the problem in question. This is where logic plays an important role. The revelation must pass through this filter in order to be finally accepted.

The phase for which we can teach ourselves is the preparation. This is the conscious phase of the work, which begins with the proposal and building an understanding of the problem. The idea would be to work effectively enough to lead to the eventual enlightenment. The work then has to fertilise our minds so that, after a certain time, creation is born. How can we do this? What can we do in order to fertilise our thoughts?

Phases of creation

Mathematical and scientific researchers coincide in naming many of the aspects involved in creative investigation: imagination, observation, experimentation, intuition, analogy, generalisation, reasoning, strategy and luck. Of all of them we have picked out six as fundamental: observation, intuition, experimentation, conjecture, analogy and verification.

Now we are going to analyse a simple phenomenon from these six different points of view to see how they contribute to our solution. The phenomenon we are going to take a look at is:

The squares of natural numbers.

Observation

Observation depends on who is carrying it out. We can only recognise what we already know. If we want to perceive something unknown in a phenomenon we must pay attention to the things that surprise us about it, things that do not impede but which allow two different observers to see different things. It also tends to be the case that an observer perceives a change in a familiar scene without being aware of what it is. In any situation, the observation does not consist of just looking

but is, rather, a mental process, the conclusion of which is usually a description or interpretation of what has been observed.

In mathematics, the most common observations lead to the identification of patterns. Which pattern can we observe in the squares of the first natural numbers?

Number	Square	
0	0^2	0
1	1^2	1
2	2^2	4
3	3^2	9
4	4^2	16
5	5^2	25

From the series 1, 2, 3, 4, 5,... we have created the series 1, 4, 9, 16, 25,... What is peculiar to the second series? It is not made up of consecutive numbers. It looks like a series of random numbers, although they have been obtained in a specific way.

In order to be able to better observe the phenomenon, let's shift our attention back to the original series. Why do we say that the numbers 1, 2, 3, 4, 5,... are consecutive? Because the difference between each number and its neighbour is always 1. Let's transfer this observation to the second series. What are the differences between the squared numbers?

Number	Square	Difference
0	0	—
1	1	$1-0=1$
2	4	$4-1=3$
3	9	$9-4=5$
4	16	$16-9=7$
5	25	$25-16=9$

Eureka! The differences between the squares are the odd numbers: 1, 3, 5, 7, 9.

SEEING PATTERNS

The observation of patterns in numerical series has its risks. The question of which number follows the 5 in the series 1, 2, 3, 4, 5,... can have several answers:

6, because it is the series of natural numbers:

1, 2, 3, 4, 5, 6,...

1, as it is a repetition of five figures:

1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 1, 2, 3, 4, 5,...

8, as the odd numbers are interleaved with the powers of 2:

$1, 2, 3, 4, 5, 8, \ldots = 1, 2^1, 3, 2^2, 5, 2^3, \ldots$

Perhaps the answer is, as Wittgenstein said, any number, as the ellipsis hides what cannot be seen and allows us to imagine whatever we wish or whatever suits us.

Intuition

Observation invites us to sense (through intuition) a pattern that can be confirmed by experimentation.

Experimentation

One of the requirements of the experimentation is to make it reproducible. This does not tend to be a problem in mathematics. Nothing impedes the reproduction of the calculations for the successive squares in the series.

Number	0	1	2	3	4	5	6	7	8	9	10	11	12
Square	0	1	4	9	16	25	36	49	64	81	100	121	144
Difference	—	1	3	5	7	9	11	13	15	17	19	21	23

Experimentation confirms the observed pattern. Calculating the differences between the first 13 natural numbers (including the 0) we get 12 odd numbers: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23.

Conjecture

Conjecture is when we spread the phenomenon to cover the entire succession of natural numbers. Thus we skip from the finite to the infinite, from the particular to the general. This is the formulation of our conjecture, a theorem:

The succession of differences of the squares of the natural numbers is the succession of odd numbers.

But, how can we confirm it? It is impossible to follow this with exhaustive experimentation to cover the entire infinity of natural numbers.

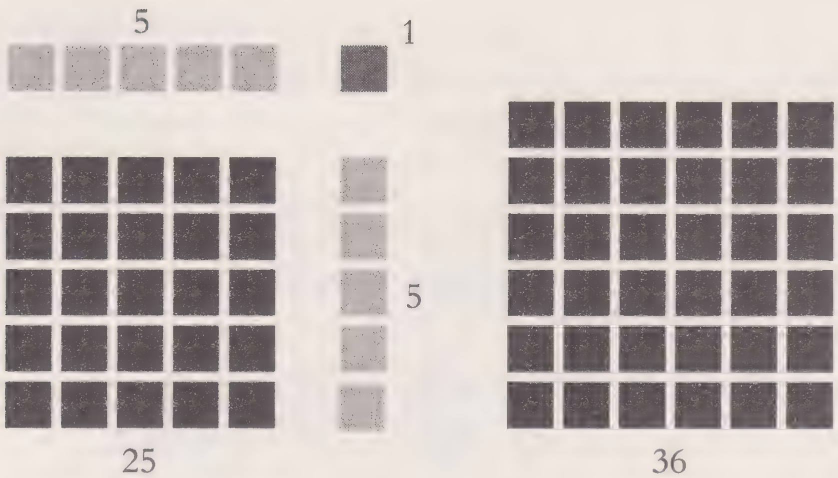
Analogy

Another question arises: do we really understand this phenomenon? The calculations tell us that the numeric reality seems to follow a pattern. But do we understand why? Do we understand why the differences between the numbers is always an odd number? With this inquiry we are not seeking to confirm or demonstrate the conjecture. What we want is to have the sensation that we understand it. Numbers and calculations speak, but they speak the language of logic. We accept the results, but despite what they say perhaps we do not see the deep undercurrent of their reality.

Analogy can help us to clarify the cause of events. What if we forget the idea of numbers and focus on the square? There is nothing stopping us from considering them to be geometric figures. In fact, the second power is termed the square because of geometric modelling. Square numbers are so-called because those quantities of units can be arranged in that way:

1 4 9

And what happens between two 'consecutive' squares? Given a square, for example, $25 = 5 \cdot 5$, what needs to be added to make it into the 'next' one, the square of $36 = 6 \cdot 6$? Let's see:



It is obtained by adding two rows of 5 units to the square of 25, plus another unit to complete the corner. That is, $36 = 25 + 2 \cdot 5 + 1$. In the same way, it can be seen that:

$$6^2 = 36 = 25 + 2 \cdot 5 + 1.$$
$$5^2 = 25 = 16 + 2 \cdot 4 + 1.$$
$$4^2 = 16 = 9 + 2 \cdot 3 + 1.$$
$$3^2 = 9 = 4 + 2 \cdot 2 + 1.$$
$$2^2 = 4 = 1 + 2 \cdot 1 + 1.$$

Here we have the key. The fact that the differences are odd is found in the corner necessary to complete the square. This analogy helps to understand the phenomenon.

Verification

We want to find definitive proof of our conjecture, which neither the numerical experimentation on each natural number nor the analogical experimentation on the square figures can provide. Both help to propose the theorem, to understand the phenomenon and to convince us of it. However, we cannot deduce the veracity of the result for all square numbers.

Perhaps it is a good idea to return to the observation to try and find a sufficiently convincing argument. Something else can be seen in the last table – the alternation between evens and odds of the natural numbers also carries over to their squares. In other words:

$$\text{even}^2 = \text{even}.$$
$$\text{odd}^2 = \text{odd}.$$

Given that the difference between even and odd is always odd:

$$\text{even} - \text{odd} = \text{odd}.$$
$$\text{odd} - \text{even} = \text{odd}.$$

We can conclude that the difference between consecutive squares will always be odd. Should we now accept the result? We are now undoubtedly absolutely convinced of it. But, is this a complete demonstration? Many will think that an algebraic version is more powerful and independent of intuition.

Let n be any natural number. Then, its consecutive number is, $n+1$. Let's carry out the squares and calculate the difference:

$$(n+1)^2 - n^2 = n^2 + 2n + 1 - n^2 = 2n + 1.$$

This number, $2n+1$, is always odd because $2n$ is even whatever n is. Then, the difference between the squares of the consecutive natural numbers is always odd. Moreover, the successive differences are also consecutive odds according to the succession of $2n+1$.

Anyone who thinks that this argument is more powerful and that it is nearing what we could take to be the definitive demonstration of the theorem should look again. It is synonymous with the analogical argument that uses the squared diagrams. The wildcard provided by n as a representation of any natural number hides the intuitive geometrical aspect by means of which we were able to understand the essence of the problem.

The result of calculating the difference between $(n+1)^2$ and n^2 is $2n+1$ shows that the phenomenon is correct. By means of the analogical (geometrical) version we have understood it and are convinced of it. As Hersh said: "One can deduce that two and two is four applying the axioms and the rules of a formal logical system, but the conviction of the result comes from collecting pebbles."

The following diagram summarises the process that should be followed and illustrates the ultimate purpose of mathematical creation: To explain a phenomenon.

$$\begin{array}{l}
\text{Phenomenon} \longrightarrow \text{Observation} \longrightarrow \text{Intuition/Conjecture} \longrightarrow \text{Experimentation} \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \downarrow \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{Verification} \longrightarrow \text{Analogy} \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \downarrow \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{Explanation} \longleftarrow \text{Analogy}
\end{array}$$

Logic does not create, but it demands

Logic has been governed by explicit axioms and rules for a long time. It is based on the human being's way of thinking. For formalists, mathematics is reduced to a series of symbols governed by those rules. But this is not the philosophical view of the mathematics championed on these pages. The truth is that arguments and verification of mathematical results are based on logic, but logic alone is not enough to produce them. Mathematical creation transcends logic. An example is the following theorem:

Every power of two is even.

Such a result may be logical, but it does not contribute anything new due to its obviousness. It cannot be considered creative. Applying the rules of logic to produce new and true results from previous ones is not a creative act. A computer could do that. A creative act implies selecting and searching for meaningful results. It corresponds to questions and interests formulated within a social and cultural context which machines lack. Even formalists constitute an environment of this type. It is not machines that produce formulations, but people. Logic is industrial; it works like a production line programmed to do what it does. Mathematics is much more than an industry. On the other hand, there are human theorems that perhaps a machine would never come up with.

We should not overlook the responsibility of creative acts either. All creation has consequences, such as when it is in response to a system of structures coherence. In

ALGEBRAIC PROPERTIES OF ORDINARY ARITHMETIC

For a set C there are two binary operations defined. We express them with the symbols '+' and '·':

- Commutative: $a+b=b+a$.
  $a \cdot b=b \cdot a$.
- Associative: $(a+b)+c=a+(b+c)$.
  $a \cdot (b \cdot c)=(a \cdot b) \cdot c$.
- Distributive of '·' in terms of '+': $a \cdot (b+c)=a \cdot b+a \cdot c$.

other words, it is the coherence of that system that inspires it. That is what happened with the rule of signs:

$$- \times - = +$$

This rule was established to preserve the coherence of the multiplication of negative integers and it is the consequence of the desire to conserve the distributive property for those numbers. In effect, the distributive property states that for any three numbers *a*, *b* and *c*, it is true that:

$$a \cdot (b+c) = a \cdot b + a \cdot c.$$

Therefore, it would be desirable that this equation were also true for $a=-1, b=1$ and $c=-1$:

$$-1 \cdot (1-1) = -1 \cdot 1 + (-1) \cdot (-1) = -1 + (-1) \cdot (-1)$$
$$-1 \cdot (1-1) = -1 \cdot 0 = 0.$$

Then, we get:

$$-1 + (-1) \cdot (-1) = 0 \Rightarrow (-1) \cdot (-1) = +1.$$

Mathematicians took a long time to understand that the 'rule of signs', together with other definitions that govern integers and fractions, could not be 'demonstrated'. We 'created' them in order to achieve operative liberty while they preserve the fundamental rules of arithmetic. We agree with Courant and Robbins when they say that the only thing which can – and should – be proved is that the commutative, associative and distributive properties are conserved based on these definitions.

So logic leads us to surprising and, sometimes, hard-to-accept results. By accepting that minus times minus gives plus, we are not only coherent with operative logic, but also with ourselves, because it was our thinking that created the laws of the logic. It is logical to accept its consequences, even though we may not like them or they may seem strange. At this point, mathematics could have rejected negative numbers, considering them to be monstrosities that impeded the development of knowledge. They could also have taken them as a sign of madness and proof of how illogical the starting points were. However, they assumed responsibilities which allowed the previous system to be expanded, achieving coherence.

Courant and Robbins stress the creative aspect of that decision, highlighting the term *created* in italics. Accepting strange results to which properties and theories lead

us in order to introduce new elements is a common form of creation in mathematics. The results appear stranger the further they stray from the most basic calculation, with which we count the rocks in a stream, or when they attack the systems established throughout time.

Living mathematically

To do mathematics means to live through mathematical experiences. And to have mathematical experiences we need the desire to understand and explain things in a certain way, according to the mathematical perspective, which goes through quite objective questioning on the aspects of the world based on quantification.

This is a new aspect in terms of Pólya's heuristics, as problems that are proposed can transcend the academic domain in which the most traditional heuristics is found.

The purpose of proposing mathematical questions regarding the phenomena that we experience, in or outside of the academic and cultural domain to which we belong, is to comprehend reality and the (social, cultural and technological) environment in which our lives are lived. There is nothing more creative than that, and we will return to this subject in Chapter 3.

Chapter 2

Big Ideas for Big Problems

Many of the great mathematical creations have been linked to great crises in the development of mathematical knowledge. Occasionally, a new creation or theorem results from a problem. At other times, it creates a conflict with the predominant ideas of the time. This is why some of the greatest mathematical creations have entailed great mental challenges. What was once considered irrational and nonsensical ended up making sense and resolved problems much more real and practical than ever imagined. The most extraordinary case of all is probably that of complex numbers: how do you accept that the square of a number is negative? How can you create rational sense for such a monster?

Is the development of mathematical knowledge linear? Although some people consider it to be so, this is debatable. The linearity of mathematical knowledge is possibly only apparent, a consequence, as is the case with axioms and theorems, of the way that mathematics has been presented and is still demonstrated to the public.

Counting

Counting consists of trying to determine the number of elements that form a group. A quick glance is enough for us to be able to be aware of small amounts, and we do not need to count to tell that groups of two, three or four units are different. However, it is difficult to tell quantities of groups formed by more than four or five things; we must count them.

Among the first forms of counting is that of associating a quantity with different parts of the human body. People from different parts of the world have used, and still use, anatomical routes to determine the number of objects in a set, what is technically called the 'cardinality' of that set. A herd of animals and a sack full of rice are sets whose cardinality is finite. Natural numbers also form a set, but of infinite cardinality. Distinguishing between two finite sets is simple: counting them is enough.

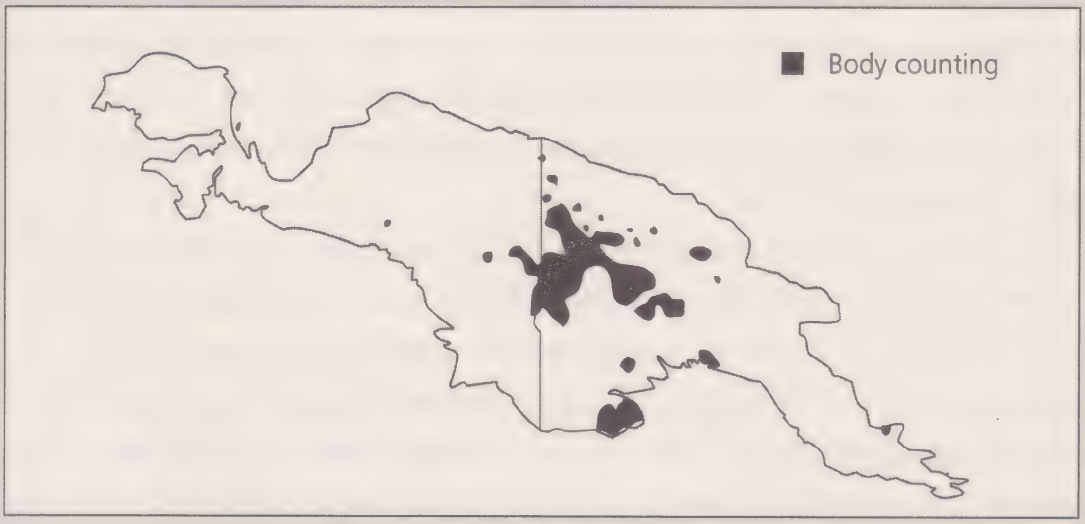
The difference is in the number that determines its cardinality. Further on we will see that the same is not true of infinite sets.

The act of counting makes sense when something finite is being counted. This occurs when we free ourselves of the tangible references and associate a symbol (verbal or written) to each quantity. As in the case of body numbers, the route is necessary for counting but, unlike them, a specific quantity is denoted by a unique symbol. This is the role of the figures 1, 2, 3, 4, 5, 6, 7, 8, 9 and 0, which we designate to the basic quantities.

But the real creation is the establishment of a base. Counting large quantities using a different symbol and a verbal term for each of them would not only become tedious, but also practically impossible, as sooner or later the terms and symbols run out, and our memory has a limit. The invention of numerical base and a positional annotation system that transforms the counting into a recurring activity is extraordinary. A base-10 positional system such as ours uses only ten symbols to represent any number, no matter how large. The verbal terms with which we refer to a quantity also

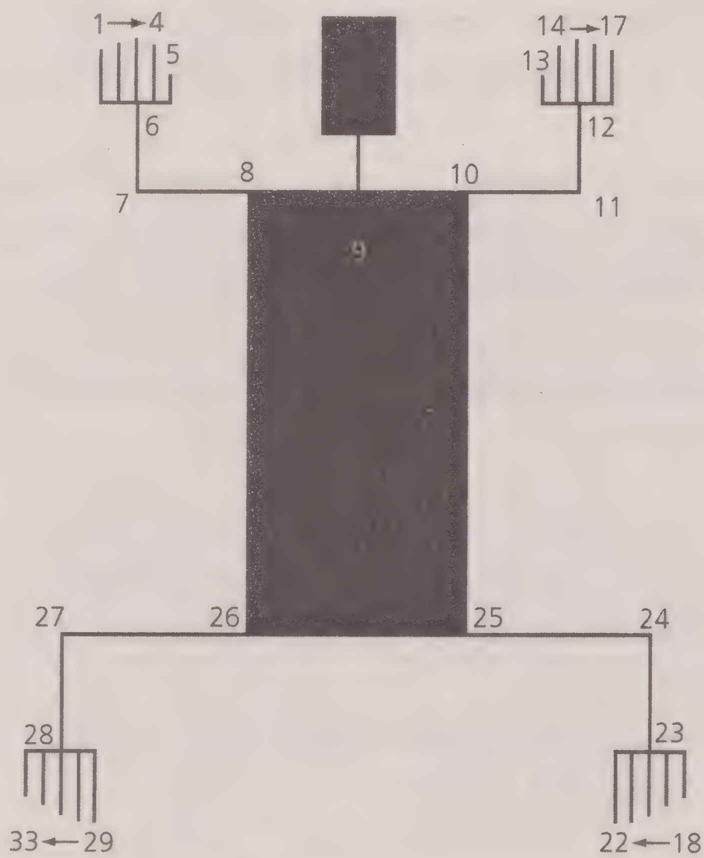
COUNTING

All cultures have developed counting systems, and most of them have done so using parts of the body which have been assigned a number. This is called 'anatomical' or 'body counting'. In 1992 researcher Glen Lean identified more than five hundred different counting systems on the island of New Guinea. The areas marked on the map indicate those where the counting is anatomical.



BODY COUNT

An example of body counting used by the islanders of the Torres Strait (which separates Australia from New Guinea), according to George Ifrah (1994). Note the asymmetry of the process with respect to the human body. The route covers the extremities and their articulations in a circular path.



follow the base and the positional system; there are 17 of them. We call them 0, 1, 2, ..., 12 and those for 100, 1,000, 1,000,000. All the rest are composed of these.

Even so, in practice the tendency is to use strategies that make counting without errors easier and instruments that help us do it. The risk of making a mistake when counting increases with the quantity being counted. So we tend to count in sets, 2 by 2, 5 by 5, in tens or dozens.

Why is it more comfortable to count in 2s than in 3s or 7s? To count in 2s all we have to do is repeat the sequence 2, 4, 6, 8, 10, adding a one to the left after each cycle, in other words, a ten:

0	2	4	6	8
10	12	14	16	18
20	22	24	26	28
30	32	34	36	...

It is not worth counting in 4s or in 8s, as 4 and 8 are multiples of 2 and, although they give the same endings, they occur in a disorderly fashion in terms of counting in evens.

4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, ...
8, 16, 24, 32, 40, 48, 56, 64, 72, 80, 88, 96, 104, ...

It is much worse counting in 3s, in 7s or in 9s. The endings do not fit a pattern that is easy to memorise:

3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, ...
7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 91, 98, ...
9, 18, 27, 36, 45, 54, 63, 72, 81, 90, 99, 108, 117, ...

If counting in 3s is uncomfortable, doing so in 6s is too, as it involves the sequential disorder of the even endings:

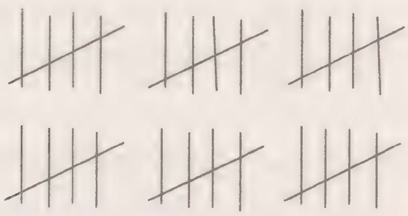
6, 12, 18, 24, 30, 36, 42, 48, 54, 60, 66, 72, 78, 84, ...

In contrast, it is easy to count in 5s or 10s:

5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, ...
10, 20, 30, 40, 50, 60, 70, 80, 90, 100, ...

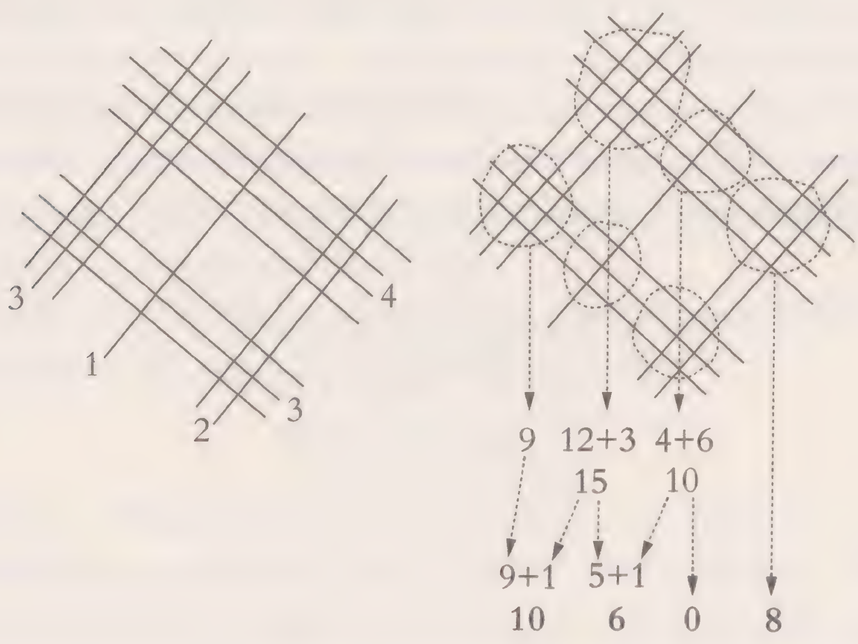
But this tends to be done after having organised the elements that need counting into groups of five or ten. In the first case, a one is added to the left (10) in each cycle (formed by a 0 and a 5). In the second, it is simply a case of counting in units and putting a 0 after them.

When it comes to counting numerous quantities, the best thing to do is distribute them into a rectangle. A calculation gives the result without having to exhaustively count everything at once.



$\Rightarrow 5 \cdot 3 \cdot 2 = 30$

The Mayan multiplication system is based on this principle. In order to multiply 312 by 34, groups separated by parallel lines are used to represent the hundreds, the tens and the units of each number. The lines of the second are distributed in such a way that they cut all those of the first. You then count the corresponding intersections. It is a visual representation of common multiplication.



However, this procedure becomes impractical with large figures or numbers, as the intersections end up being too numerous.

And what happens when we want to count infinite quantities? In common language we all use the term 'infinite' to stress the huge magnitude of something, be it a seemingly never-ending process or an enormous object. Contrary to common belief, there is not only one infinity. In mathematics, there are at least two types of infinity. One is the usual infinity derived from natural numbers, with which we count

or number objects: 1, 2, 3, 4, ... The other is uncountable infinity with which we count the points on a line.

With infinities come paradoxes. For example, it is difficult to believe that natural numbers have the same cardinality (infinite) as a subset of them, such as the set of even numbers. How is this possible if there are twice as many natural numbers as even numbers? Necessarily, there are twice the number of the former as there are of the latter, but the double of infinity is infinity.

The issue does not raise questions if we say what is understood by an infinite set. It is said that a set is numerable – that it can be counted – if its elements can be aligned with the natural numbers. So, it is clear that the even numbers can be counted and that the relationship established by the count determines their cardinality.

2	4	6	8	10	12	14	16	18	20	22	24	26	...
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	...
1	2	3	4	5	6	7	8	9	10	11	12	13	...

It is perhaps more surprising that all rational numbers have the same cardinality as the natural numbers. They are counted by expressing them first as a fraction, distributing them suitably and establishing the order of counting.

1/1 1/2 1/3 1/4 ...
2/1 2/2 2/3 2/4 ...
3/1 3/2 3/3 3/4 ...
4/1 4/2 4/3 4/4 ...
... ...

That of natural numbers is the 'basic' infinite cardinality, and it is represented by the sign $\aleph_0$. The cardinality of an infinite set of numbers which, unlike the previous set, is not numerable – it is not countable with integer numbers – is represented with the symbol $\aleph_1$. Another way of looking at it is to say that its elements cannot be aligned with the natural numbers. Its infinity is of another nature.

A very familiar example of a set of non-numerable numbers is that of real numbers between 0 and 1. (This includes irrational numbers, such as the square root of 2, which are not a quotient of two integer numbers.)

The proof shown below, one of the most amazing in mathematics, was invented by the great Georg Cantor.

Imagine that we had managed to count all the real numbers between 0 and 1. Then we could put them in a list such as the following:

1	0.**0**37563856636663...
2	0.9**1**9688568847383...
3	0.15**5**382300008691...
4	0.000**0**00033433002...
5	0.9999**9**5885994382...
6	0.10100**1**000100001...
7	0.774647**7**46477464...
...	

Then an infinite real number 0. ... can be constructed that will not be on the list. It is constructed as follows: if the first decimal of the first number on the list is 1, we write 0; if it is not 1, we write 1. Following to this premise and according to the above list, our number would start with 0.1...

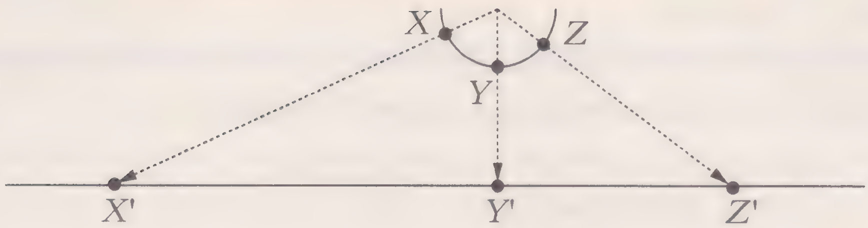
We then look at the second decimal of the second number on the list and apply the same argument as before. If that decimal is 1, we write 0; otherwise, we write 1. Our number now has two decimal points: 0.10...

We repeat the same reasoning again in order to add more decimal places to the number. According to the above list, our number would be

$$\Psi = 0.1011101...$$

This number is different from all those on the list, as it differs in at least one decimal. Therefore, it is not part of it. In fact, it is the list itself that creates it. Therefore, a full list is impossible, and the real numbers between 0 and 1 cannot be counted.

This trick is known as Cantor's 'diagonal argument', and it shows us that the nature of the infinity of any real numbers is not the same as that of natural numbers. Some paradoxes arise here too. For example, although they differ in length, there are as many points on a semicircle, regardless of its size, as all of those on a line. A very simple graphic proves this apparent nonsense. If we draw all the radii from the centre of the circle in question which pass through the semicircle, we will establish a one-to-one relationship between the points on the semicircle (X, Y, Z,...) and those of the line (X', Y', Z',...) as shown overleaf.



Unnatural powers

People approach new ideas carrying cultural baggage. When new ideas do not fit with the previous patterns it is essential to make a change of focus. In the process of assimilation the student may find a conflict between their own internal reasoning and external reasoning.

This is what happens with the concept of negative, decimal or irrational exponents, which challenge the established understanding of basic operations such as multiplying and dividing.

Raising a number to a power is to multiply it by itself as many times as indicated by the power:

$$3^4 = 3 \cdot 3 \cdot 3 \cdot 3.$$

When multiplying powers, the exponents are added, and when dividing, they are subtracted:

$$2^3 \cdot 2^5 = (2 \cdot 2 \cdot 2) \cdot (2 \cdot 2 \cdot 2 \cdot 2 \cdot 2) = 2^8.$$

$$\frac{2^7}{2^4} = \frac{2 \cdot 2 \cdot 2 \cdot \cancel{2} \cdot \cancel{2} \cdot \cancel{2} \cdot \cancel{2}}{\cancel{2} \cdot \cancel{2} \cdot \cancel{2} \cdot \cancel{2}} = 2^3.$$

But if we divide powers with the same exponent, for example 2^3 by 2^3, we get a surprising result. On the one hand the result is 1, as $8/8 = 1$. But according to the above rule, the exponents should be subtracted and we get:

$$\frac{2^3}{2^3} = 2^{3-3} = 2^0.$$

This means that coherence is only conserved if $2^0 = 1$. How is it possible that a number multiplied by itself zero times equals 1? Things do not end there. If, in a division of powers, the denominator is the greater of the exponents, the result is a power with a negative exponent.

$$\left.\begin{array}{l} \dfrac{2^2}{2^4} = \dfrac{2\cdot 2}{2\cdot 2\cdot 2\cdot 2} = \dfrac{1}{2^2} \\ \dfrac{2^2}{2^4} = 2^{2-4} = 2^{-2} \end{array}\right\} \Leftrightarrow \frac{1}{2^2} = 2^{-2}.$$

Elevating a number to another used to be multiplying it various times. Now we have created operations and expressions that clash with our way of understanding previous ones. Elevating a number to a negative means dividing one by the number as many times as indicated by the exponent. Is this logical? Does it make sense? It is logical, of course, but the sense needs to be changed. The conception of the exponent as an indicator of the number of factors of the product need not be modified. Also, a negative exponent means a positive exponent in the denominator of a fraction. Here we have the transition from the negative exponents to the positive ones:

$$a^{-2} = \frac{1}{a^2}, \;\; a^{-1} = \frac{1}{a}, \;\; a^0 = 1, \;\; a^1 = a, \;\; a^2 = a\cdot a.$$

Fractional or decimal exponents are governed by a similar process. If the square root of a number is elevated to its square, the result is the original number:

$$\left(\sqrt{a}\right)^2 = a.$$

Which exponent corresponds to a number such as the square root of *a*?

$$\left.\begin{array}{l} \left(\sqrt{a}\right)^2 = a \\ \left(\sqrt{a}\right)^2 = \left(a^x\right)^2 = a^{2x} \end{array}\right\} \Leftrightarrow 2x = 1 \Leftrightarrow \sqrt{a} = a^{\frac{1}{2}}.$$

Having come this far, why not consider the meaning of expressions such as this?

$$2^{\pi},\ 2^{\sqrt{2}}.$$

Its meaning comes from the fact that all irrational numbers (those which are not fractions of two integers) are the limit of a succession of finite real numbers, as is the case with the square root of 2 and the number π:

$$1;\ 1.4;\ 1.41;\ 1.414;\ 1.4142;\ 1.41421, \ldots \sqrt{2}.$$
$$3;\ 3.1;\ 3.14;\ 3.141;\ 3.1415;\ 3.14159, \ldots \pi.$$

Given that we know what it means to elevate a number to a decimal, we can now define the power of an irrational exponent:

$$2^{\sqrt{2}} = \text{limit of } \{2^{1};\ 2^{1.4};\ 2^{1.41};\ 2^{1.414};\ 2^{1.4142};\ \ldots\}.$$

We have come so far from the original definition of powers! Here we have extraordinary mathematical creations, as with the basic operations we have created new ones that we have given new meanings. Their meaning clashes with previous conceptions, but they are guided by the logic of calculation and the desire for coherence. Powers started off with natural exponents. But now, those with natural exponents are perceived as a mere individual case of a much larger construction. Exponents can be negative, decimal and even irrational.

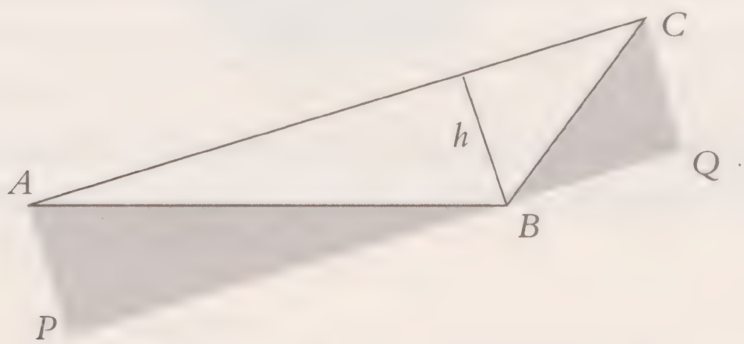
Assimilating a creation entails a change of perspective. We can no longer think that a number elevated to another consists in multiplying it by itself as many times as the exponent dictates, because it makes no sense to multiply a number by itself 0.12 times or π times. We have freed ourselves of that basic point of view. The basic view served as a springboard for elaborating a much broader and general creation of which the original is seen as a mere individual case. Creation has changed us.

From the area of a rectangle to that of any figure

The line segment and the triangle are fundamental figures in mathematics and in all human knowledge. The former only has one magnitude: length. In fact, and given that there is no tangible object capable of representing it faithfully, it can be said that the segment is made of length. The triangle, on the other hand, has length (a perimeter) and area, the amount of space enclosed by its three sides.

Since ancient times, the calculation of areas has been a fundamental problem for us humans. The most popular legend on the origin of mathematics takes place on the Nile, and even its purpose is the measurement of the area of the flooded areas of its floodplains.

Given a rectangle of sides a and b the area A enclosed by its four sides is defined as the product of the length and its width: $A = a \cdot b$. Given that all triangles are half a rectangle, their area is half the area of that rectangle. Indeed, in the following figure the area of triangle ABC is half of rectangle $APQC$, the base of which is side AC of the triangle and the width, height h over AC:



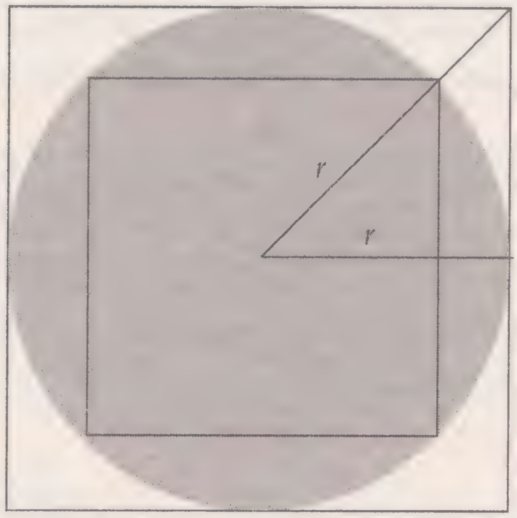
Therefore, the area of the triangle is half the product of the base multiplied by the height.

$$A = \frac{1}{2} \cdot AC \cdot h.$$

Any polygonal flat figure can be broken down into a series of triangles. Calculating its area translates into calculating the sum of the areas of the triangles which it is composed of. But, what happens if the profile of the shape is not made of straight lines, but curves.

The simplest curved shape is the circle. The problem of the calculation of the area of a circle is very old, and the question of whether it is possible to construct a square with a ruler and compass with an area that is equal to a given circle is one of the three classic problems of geometry.

Is there a relationship between the area of a circle and that of a square? A first attempt to find the area of a circle of radius r could be to inscribe a square in it and circumscribe it into another one:



The area of circle A_C is somewhere between the areas of a square with diagonals of $2r$ and another of sides of $2r$. Their average provides an estimation of area A of the circle:

$$2r^2 < A_C < 4r^2 \quad \Rightarrow \quad A_C \approx \frac{2r^2 + 4r^2}{2} = 3r^2.$$

We now know that the result is incorrect, given that the area of the circle is not $3r^2$, but πr^2. But in ancient Egypt they used a value of 3 as the ratio between the circumference and the diameter of the circle, although it can easily be proved by a complete rotation of a wheel of radius r, that the length it describes exceeds three times the diameter. But it is not the value of π that we are now concerned with, but how to get from the area of a rectangle or a triangle to the area of a circle.

We can try to make do with an equilateral triangle which we have made with a square and inscribe and circumscribe the circle, but then the problem becomes over-complicated, and the result is no better than the previous one. Following this model we see that inscribing and circumscribing the circle between two regular polygons of multiple sides gives a much better approximation. The result will be more precise the more sides the regular polygons have.

Finally, in the limiting case, if it exists, we would have two polygons with infinite sides whose areas should coincide with the area of the circle. Then, it is sufficient just to work with the inscribed polygon or the circumscribed one because in the limiting situation they become the same for our purposes.

This is what Archimedes did. He converted the calculation into a process. Instead of taking polygons with n sides, what he did was to start with a regular hexagon and

then double the number of sides several times. Thus he obtained the measurements corresponding to the polygon with 96 sides and an excellent approximation of the number π and of the area of the circle:

$$\frac{223}{71}r^2 < A_C < \frac{22}{7}r^2 \quad \Rightarrow \quad A_C \approx \frac{3123}{994}r^2 = 3.14185\ldots r^2.$$

Credit is due to Archimedes in this case, not for carrying out the necessary tedious calculations, but for two other aspects. On the one hand, for saving himself the trouble of doing too many calculations by demonstrating that the perimeters and inscribed and circumscribed areas of one stage of the process can be used to calculate those of the next by means of the harmonic and geometric measurements. On the other, for creating an iterative method which gives a more precise result at each step. This means creating a path to infinity. The end will never be reached, but we can discover what is along the way.

ARCHIMEDES IN THE 21ST CENTURY

Trigonometry and current technology allow a variant of the Archimedean procedure to be carried out creating a recurrent method based on regular polygons with 2^n sides. The area of the 2^n-agon inscribed in the circle with a radius of 1 is:

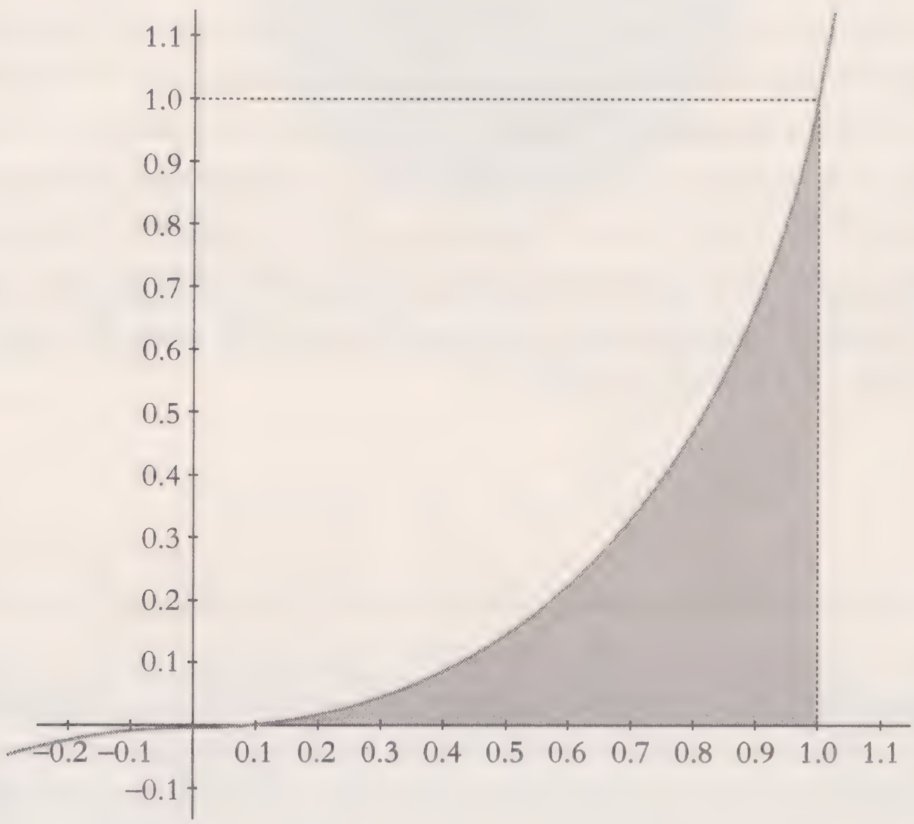
$$A_n = 2^{n-1}\mathrm{sen}\left(\frac{\pi}{2^{n-1}}\right).$$

With trigonometric analysis we can see that the law of formation lies in the angles of the type $\pi/(2^n)$. This law leads to the area of circle A_C:

$$A_C = \lim_{n\to\infty} A_n = \lim_{n\to\infty} 2^{n+1}\sqrt{\underbrace{2-\sqrt{2+\sqrt{2+\ldots}}}_{n}}.$$

Software available today allows the calculation of the area of the polygon with $1{,}024=2^{10}$ sides with the previous formula and shows that it is very close to area expected: $A_{1024}=3.1415923\ldots$

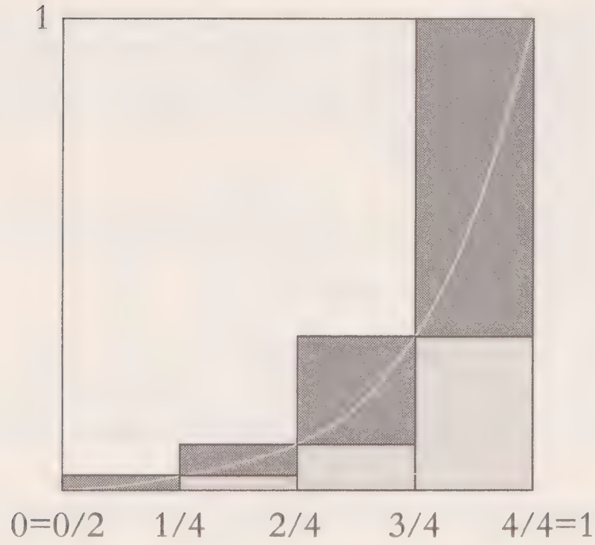
The circle is the just simplest of the curved shapes. How do we calculate the area of any other shape? Knowing the formula corresponding to the curved part of a shape, mathematicians resolve the problem by means of a method similar to that employed by Archimedes. Let's say we want to determine the area enclosed between the abscissa axis and curve $y=x^3$ from the origin at coordinates (0,0) to point (1,0), the shaded area of the image:



A first approximation of that area would be that of the right-angled triangle with vertices (0,0), (1,0) and (1,1), which is 1/2. However, this produces an extraordinarily large overestimation.

The method described below, called the 'method of exhaustion', was already used by Archimedes more than 2,000 years ago to estimate the area of a segment of a parabola. Our first approximation was based on the area of a triangle similar to that of a shape. Now use the area of a rectangle.

We divide interval [0,1] into four regular intervals and place two rectangles in each of them, one whose height is based on the left-hand side and the other based on the right-hand side. As $f(0) = 0^3 = 0$, the height of the first lower rectangle is 0:

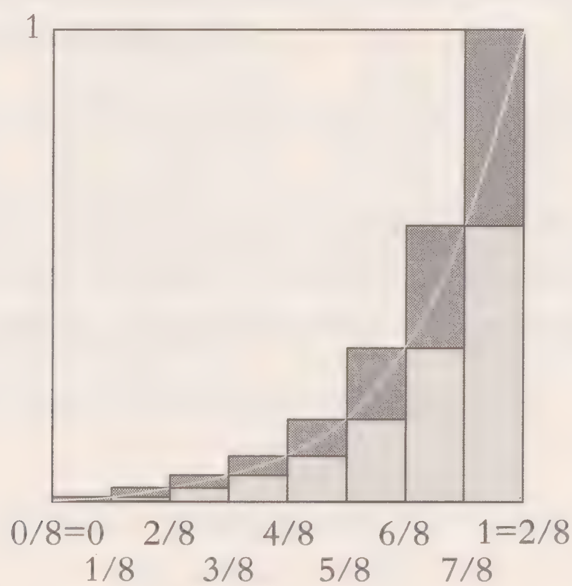


The area we are looking for, A, will be found between the lower A_L (light grey) and upper A_U (dark grey) areas. In short, it will be greater than the first and lesser than the second. We are going to calculate them both taking into account that all the rectangles have the same base, 1/4, and that only their heights change.

$$A > A_I = \frac{1}{4}\left[(0)^3 + \left(\frac{1}{4}\right)^3 + \left(\frac{2}{4}\right)^3 + \left(\frac{3}{4}\right)^3\right] = \frac{1^3 + 2^3 + 3^3}{4^4} = 0.140625.$$

$$A < A_S = \frac{1}{4}\left[\left(\frac{1}{4}\right)^3 + \left(\frac{2}{4}\right)^3 + \left(\frac{3}{4}\right)^3 + \left(\frac{4}{4}\right)^3\right] = \frac{1^3 + 2^3 + 3^3 + 4^3}{4^4} = 0.390625.$$

The average of both areas is: $A \approx (A_I + A_S)/2 = 0.265625$. We can improve this estimation by refining the partition:



Now each rectangle has a base of 1/8. Again, the sum of the areas of the dark rectangles (A_U) is greater than the area we are looking for. And this is greater than the areas of the light rectangles (A_L):

$$A > A_I = \frac{1}{8}\left[\left(\frac{0}{8}\right)^3 + \left(\frac{1}{8}\right)^3 + \ldots + \left(\frac{7}{8}\right)^3\right] = \frac{1^3 + 2^3 + \ldots + 7^3}{8^4} = 0.1914\ldots$$

$$A < A_S = \frac{1}{8}\left[\left(\frac{1}{8}\right)^3 + \left(\frac{2}{8}\right)^3 + \ldots + \left(\frac{8}{8}\right)^3\right] = \frac{1^3 + 2^3 + \ldots + 8^3}{8^4} = 0.3164\ldots$$

The average is:

$$A \approx \frac{1}{2}\left(A_S + A_I\right) = 0.2539\ldots$$

If the process is continued by creating increasingly narrower partitions of the interval [0,1], in other words, with a far greater number of parts, we can see that in the extreme case the number of parts and rectangles would be infinite, and that the sum of its areas would provide precisely the area below the curve. The question is how to obtain the definitive area of those infinite rectangles. With what we have done so far we can say that the value we are looking for must be around 0.25, as that is what is indicated by the results obtained: 0.2656... and 0.2539...

In order to provide an accurate answer to the problem, let's have a look at how we arrived at the two previous values. Whether there are eight, one hundred, one thousand or *n* rectangles, the calculation of the sum of their areas will be carried out in the same way. The value of area A_U for a partition of [0,1] into *n* equal parts, will be the following:

$$A_S = \frac{1}{n}\left[\left(\frac{1}{n}\right)^3 + \left(\frac{2}{n}\right)^3 + \ldots + \left(\frac{n}{n}\right)^3\right] = \frac{1^3 + 2^3 + \ldots + n^3}{n^4}.$$

Therefore, the problem is reduced by finding the number towards which the final quotient converges when *n* becomes so big that it tends towards infinity. Let's have a look at the results of the numerator, the sum of the cubes of the natural numbers:

$$1^3 = 1$$
$$1^3 + 2^3 = 9$$
$$1^3 + 2^3 + 3^3 = 36$$
$$1^3 + 2^3 + 3^3 + 4^3 = 100.$$

The results are 1, 9, 36, 100, ..., or, the squares of 1, 3, 6, 10, ... It seems that by adding the cubes of the natural numbers we get the squares of... which numbers? What series is formed by 1, 3, 6, 10, ...? Let's take a look at the following:

$$1=1$$
$$1+2=3$$
$$1+2+3=6$$
$$1+2+3+4=10.$$

We seem to have hit upon a theorem.

The sum of the cubes of the first n *natural numbers is the square of their sum.*

From an experimental point of view, the validity of that statement can be verified for very large quantities of the first natural numbers. It is something a computer can do in an instant. But the experimental observation of specific results and the induction of a general principle based on them, which is the way in which physicians and biologists normally function, are not sufficient for mathematicians. They need to prove the veracity of the phenomenon observed in each and every possible case.

So, how can the veracity of the theorem be checked for each and every possible case? We will start by calculating the first n natural numbers with a process similar to that already employed by the illustrious German mathematician Carl Friedrich Gauss when he was counting at just the age of ten. Biographers recount a story of a professor, who, in order to keep his students busy, ordered them to calculate the sum of the natural numbers from 1 to 100. Among the students was Gauss, who after a few seconds, and to the surprise of the tutor, handed in a slate with the correct answer. The young genius had arranged the two series one above the other and had added them by columns. Like this:

$$\begin{array}{ccccccc} 1 & +2 & +3 & +\ldots & +98 & +99 & +100 \\ 100 & +99 & +98 & +\ldots & +3 & +2 & +1 \\ \hline 101 & +101 & +101 & +\ldots & +101 & +101 & +101 \end{array}$$

The sum of the lower row is $100 \cdot 101 = 10{,}100$, which is two times the requested amount. The correct value of the sum is, therefore:

$$1+2+3+\ldots+99+100=\frac{10{,}100}{2}=5{,}050.$$

Applying this strategy to our case and in a more general manner:

$$\begin{array}{ccccccc} 1 & +2 & +3 & +\ldots & +(n-2) & +(n-1) & +n \\ n & +(n-2) & +(n-1) & +\ldots & +3 & +2 & +1 \\ \hline \end{array}$$
$$(n+1)+(n+1)+(n+1)+\ldots+(n+1)+(n+1)+(n+1)=n\cdot(n+1)$$

We can see that the formula for observing the first n natural numbers is:

$$1+2+3+\ldots+n=\frac{n(n+1)}{2}. \quad [\star]$$

Going back to the theorem, we write this result in the new formula:

$$1^3+2^3+\ldots+n^3=\left(1+2+3+\ldots+n\right)^2=\left(\frac{n\left(n+1\right)}{2}\right)^2=\frac{n^2\left(n+1\right)^2}{4}. \quad [\star\star]$$

We now have two formulae on the first n natural numbers. The mathematician does not attempt to carry out exhaustive observations on infinite sets, such as that of the natural numbers, given that it is clearly impossible. What he does is to come up with rigorous strategies which avoid that problem. He thinks in the following way: "Very good, we know that the formula is true up to the nth natural number. Given that all natural numbers are obtained by adding a one to the preceding number, from the veracity of the formula for the nth number, I can demonstrate its compliance for the following natural number. In other words, if the formula is valid for n, I can prove that the formula for $n+1$ is also valid, and so I will have proved the certainty of the general formula for all natural numbers".

This is what we are going to do. First, we will prove that the validity of [⋆] for n also applies to $n+1$. Then we will do the same with [⋆⋆]. We will thus demonstrate that:

$$1+2+3+\ldots+n=\frac{n\left(n+1\right)}{2}\Rightarrow 1+2+3+\ldots+n+\left(n+1\right)=\frac{\left(n+1\right)\left(n+2\right)}{2}.$$

It only has to be proved that the difference between the two expressions on the left-hand side of the equation, whose value is $n+1$, is the same as the difference between the expressions on the right-hand side:

$$n+1=\frac{(n+1)(n+2)}{2}-\frac{n(n+1)}{2}.$$

All that is left is to calculate the right-hand side of the equations to see that this is indeed true. The same applies to [★★]. We can now conclude our area problem:

$$A_S=\frac{1^3+2^3+...+n^3}{n^4}=\frac{(1+2+3+...+n)^2}{n^4}=$$

$$=\frac{1}{n^4}\left(\frac{n(n+1)}{2}\right)^2=\frac{n^4+2n^3+n^2}{4n^4}=\frac{1}{4}+\frac{1}{2n}+\frac{1}{4n^2}\xrightarrow[n\to\infty]{}\frac{1}{4}+0+0=\frac{1}{4}.$$

The last step is justified because as n increases, quotients $1/2n$ and $1/4n^2$ get increasingly smaller: *At the theoretical limit of* n*'s growth – infinity – the value of both quotients is 0.* Consequently, the area below curve $y=x^3$ is $1/4=0.25$.

From this whole process, the most outstanding mathematical creation is that of inscribing approximate enclosures which, as they are rectangular, have an area that is easy to calculate on the interior of the curved enclosure, whose area we are trying to calculate. As we refine the partition we are 'completing' that enclosure with greater precision. Given that the values in the areas of the approximate enclosures approach, 'at the limit', a specific number – and we can demonstrate that this is indeed the case, we can attribute that specific value to the area of the curved enclosure. From a geometric parallel we move on to another numeric one, in reverse. We are resolving a problem that is more approachable than the one proposed and we use it as a platform from which we build a solution.

The use of limits in the resolution of problems is one of the greatest mathematical creations in history. At the end of the 17th century, Newton and Leibniz used it to develop infinitesimal calculus. A century and a half later, Frenchman Cauchy and the German Wierstrass refined the concept of 'limit' for the case of continuous functions such as the one we just looked at.

Quantifying change

The invention of infinitesimal calculus enormously favoured the development of mathematics, of physics and of science in general. Newton and Leibniz developed it – according to historians – independently. Basically, Newton and Leibniz's contribution

is the answer to the following question: how can we quantify the instantaneous change of a phenomenon?

Quantifying the change between two instants does not entail any problems. All we have to do is subtract the values at each of those two moments. For example, if the function described by the phenomenon is $f(t) = t^2$, where t represents the time in seconds, the magnitude of the change between the instants $t = 0$ and $t = 1.5$ will be 2.25 units:

$$f(1.5) - f(0) = 1.5^2 - 0^2 = 2.25.$$

But this means of quantifying change is of no use, because in a much shorter interval of time, such as that between $t = 4.77$ and $t = 5$, the change is practically the same:

$$f(5) - f(4.77) = 5^2 - 4.77^2 \approx 2.25.$$

The value obtained from the difference does not really indicate what is happening. What we want is a number that indicates the magnitude of the change to contemplate the interval in which it is produced. The same change in a short interval of time is more significant than in a very long period of time. Therefore, the magnitude of change should take into account the proportion with respect to the time in which it occurs, what is called 'rate of change'.

$$\frac{f(1.5) - f(0)}{1.5 - 0} = \frac{1.5^2 - 0^2}{1.5} = 1.5.$$

$$\frac{f(5) - f(4.77)}{5 - 4.77} = \frac{5^2 - 4.77^2}{0.23} = 9.77.$$

This is an improvement. The rate of change reflects what we wanted, as 9.77 is much greater than 1.5. However, the problem was to find a way to answer the question in an instant, not in an interval. How can we quantify the change at an instant of, for example, 1 second?

The mathematical strategy consists of calculating rates of change for intervals, which are increasingly smaller and closer to the instant $t = 1$ and seeing which numeric value the results approach:

Enter $t=1$ y $t=5$: $\dfrac{5^2-1^2}{5-1}=\dfrac{24}{4}=6.$

Enter $t=1$ y $t=2$: $\dfrac{2^2-1^2}{2-1}=\dfrac{3}{1}=3.$

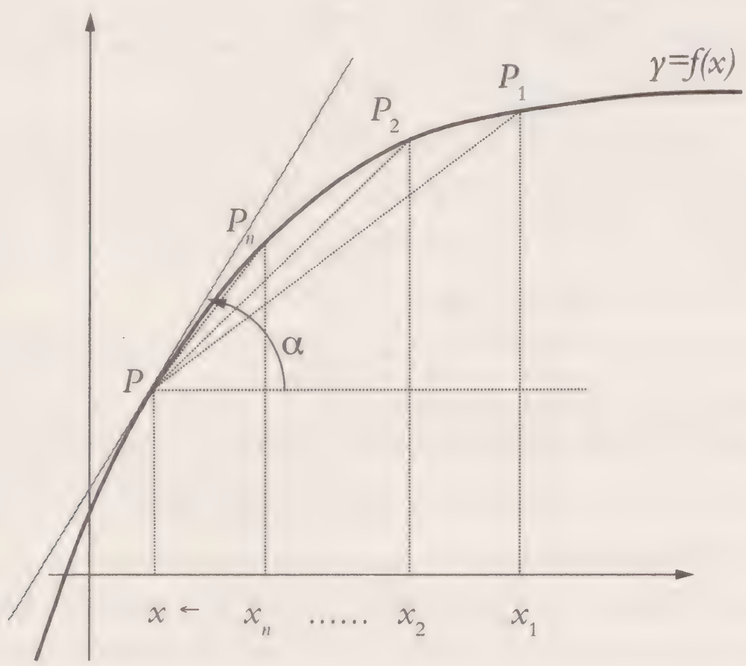
Enter $t=1$ y $t=1.05$: $\dfrac{1.05^2-1^2}{1.05-1}=\dfrac{0.1025}{0.05}=2.05.$

Enter $t=1$ y $t=1.003$: $\dfrac{1.003^2-1^2}{1.003-1}=\dfrac{0.006009}{0.003}=2.0003.$

We can clearly see that the values obtained approach the value of 2. This will be the value of change we are going to assign to $t=1$. This is given the name of 'instantaneous rate of change'.

From a graphical point of view, the rate of variation equals the value of the slope, or gradient, of the curve at a point on it, given that such slope is specifically calculated by dividing the two values of the function at each end of the interval by its length.

As values $x_1, x_2, x_3, \ldots$, approach x, points $P_1, P_2, P_3, \ldots$ approach P (see the diagram below). So we attribute point P with a slope of a value equal to that approached by the slopes of each of the preceding points.



A theorem that creates monsters

Pythagoras, the most famous of the mathematicians, created the most popular theorem, and one of the most used worldwide day-to-day, in and outside of the academic field. The exercises that are made with it in secondary education are often fun puzzles. It is all very different from the work offered by Euclid, who gives a demonstration that is also based on the calculation of areas. As stated by the theorem, this makes reference to the areas of squares constructed on the sides of a right-angled triangle. But, as with another famous theorem, that of Thales, the areas are only of use for demonstration. The theorem is restricted to calculating lengths.

The normal implication from the theorem is:

If a, b, c *are the sides and the hypotenuse of a right-angled triangle, then* $a^2+b^2=c^2$.

The reciprocal, and sufficient, implication is practically never presented:

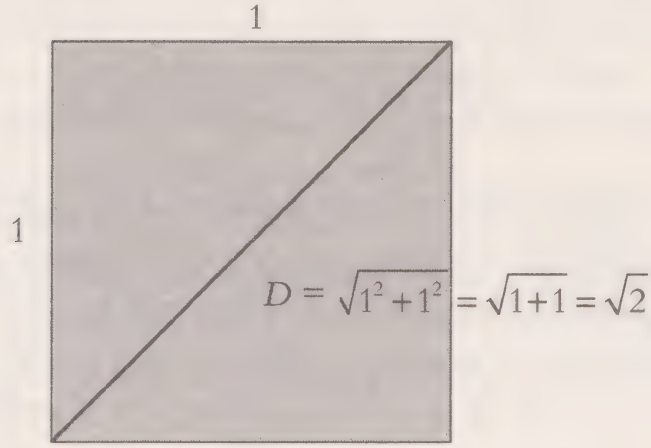
If $a^2+b^2=c^2$, *then* a, b, c *are the sides and hypotenuse of a right-angled triangle.*

This implication has enormous practical significance, for example, building a right-angled structure such as the wall of a building. This method was used by the Egyptians, who employed standard triangles with sides 3, 4 and 5 units. This relationship between the sides of a right-angled triangle has been discovered in all corners of the world in very different historical periods.

Long before Pythagoras, in ancient Egypt and Mesopotamia, sets of three numbers were known that defined a right-angled triangle. They would later be called Pythagorean triples. The square of the larger integer can be split into the sum of the squares of the two other integers.

What was unknown in any case was a proof of the phenomenon. We find numeric curiosities, such as, for example, $5^2+12^2=13^2$, but without knowing what causes them or what implications they have, it is little more than trivia. The actual demonstration of the theorem by the Greek Pythagoras is linked to the first great crisis of mathematical knowledge.

The motto of Pythagorean philosophy was 'All is number', and it was based around a numeric mysticism that all things were related by the proportions between the natural numbers. Then there came a surprise when applying the Pythagoras theorem to the diagonal of a square:



According to Pythagorean thought, that length D (the square root of two) should have been a length that was commensurable with the side of the square, that is, an integer fraction of the side. Dividing the side of the square into enough parts, let's say one million parts, the length of the diagonal should correspond to an exact number of those parts. Perhaps 1,414,213? No, because there is no way of expressing the square root of two as a quotient of natural numbers, which impeded the search for a simple calculation of the side of a square from its diagonal.

The theorem had created a monster that contradicted its most deeply-rooted conceptions. Not everything was reducible to the proportion between two integers. Something as simple as the diagonal of a square was not proportional to its side. Thus, so-called incommensurable magnitudes were born. At the time there was not sufficient knowledge to show that the circumference of a circle was also incommensurable with its diameter, and that meant the number π was also incommensurable.

Let's see why the square root of 2 is not a quotient of two natural numbers. All natural numbers n can be broken down as a product of prime factors. For example:

$$12 = 2^2 \cdot 3.$$
$$315 = 3^2 \cdot 5 \cdot 7.$$

Note that the factorial decomposition of a number to the square contains all of its factors an even number of times:

$$12^2 = (2^2 \cdot 3)^2 = 2^4 \cdot 3^2.$$
$$315^2 = (3^2 \cdot 5 \cdot 7)^2 = 3^4 \cdot 5^2 \cdot 7^2.$$

For the quotient of two natural numbers m and n to be the square root of 2 then:

$$\frac{m}{n} = \sqrt{2} \Leftrightarrow \frac{m^2}{n^2} = 2 \Leftrightarrow m^2 = 2n^2.$$

And now, both the breakdown into factors of m^2 and of n^2 will have an even number of prime factors. For this reason, and independently of whether or not there are any 2s in the break-down of n^2, the quantity of 2 in the breakdown of $2n^2$ will be odd. If n^2 does not contain a 2, in $2n^2$ there will be one (the one adjoining it); if n^2 contains a 2, it will be in an even quantity and, therefore, in $2n^2$ there will be an odd number. This makes it impossible for m^2 and n^2 to be equal, as the number of 2s in one is even and, in the other, odd. Therefore, $\sqrt{2}$ cannot be the quotient of two natural numbers.

TRANSCENDENTAL NUMBERS

A polynomial is a formula in which a variable is raised to natural powers. The numbers that multiply those powers are called coefficients. For example, the following polynomial,

$$P(x) = x^5 - 4x^3 + \frac{3}{2}x^2 - 6,$$

has rational coefficients. These are: 1, –4, 3/2 and –6. It is said that a number a is the 'root of a polynomial' if substituting it into the variable gives the result of zero: $P(a)=0$. The number $a=2$ is the root of the above polynomial. A number is said to be 'transcendental' if it is not the root of any polynomial with rational coefficients. In other words, that an equation cannot be written with natural powers to which the number in question is the solution. The irrationality of $\sqrt{2}$ was already demonstrated by the Greeks. That of the number π, although suspected, took a lot longer: Lambert demonstrated it in 1761. In 1882, Lindemann proved that it was transcendental. As a result, the impossibility of squaring the circle was confirmed definitively. The number e ($e=2.71828182845904...$) takes its name from the initial of one of the greatest mathematicians in history, Swiss Leonhard Euler (1707–1783). Like π, e is irrational. And transcendental.

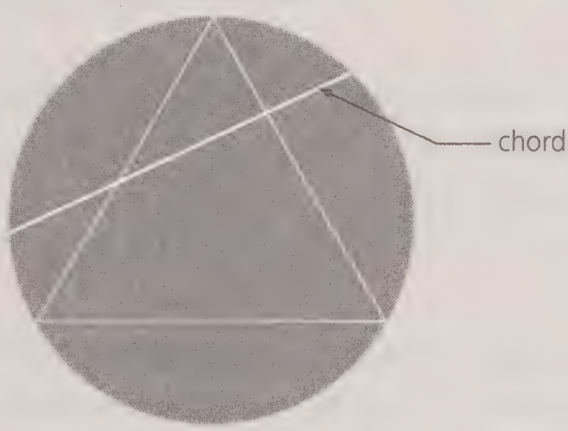
Natural numbers are so familiar to our thinking that they are considered to be more divine creation than human creation. It would be said that something so perfect cannot have faults and that any theorem based on natural numbers has to end up being demonstrated, affirmatively or negatively. It has to be possible to assert any affirmation established on such system as true or false.

However, the mathematician Kurt Gödel (1906–1978) demonstrated that this is not so, and that natural numbers can produce undecidable theorems, that is, those from which one of the two options cannot be concluded. Gödel's so-called 'theorem of incompleteness' stated that natural numbers also produce paradoxes.

PARADOXES

A paradox is the manifestation of a contradiction whose incompatible terms are true. Recurrence in language creates paradoxes, such as in the first two cases. The third is an extraordinary mathematical problem with three different solutions.

1. A barber in a barber shop only shaves those men who do not shave themselves. Who shaves the barber?
2. An adjective is said to be tautological when it defines itself. For example, brief is tautological, as it is a brief adjective. Adjectives which are not tautological are called heterological. What kind of adjective is 'heterological'?
3. The Bertrand Paradox. A chord is drawn at random in a circle. What is the probability that its length exceeds the length of the side of the equilateral triangle inscribed inside the circle. The probability can be calculated in three different ways, which lead to three different results: 1/2, 1/3 and 1/4.



How to create and tame a monster

Finding meaning and significance for the products of mathematics has always been a focus of creativity. There are many simple equations that are said not to have a solution because the number that would resolve them does not make sense in the normal numeral system.

In the field of natural numbers, in which whole numbers are used for counting things, the following equation has no solution, as the only possible solution is not a natural number:

$$2x = 1.$$

But it does have one in decimal numbers:

$$2x = 1 \iff x = \frac{1}{2} = 0.5.$$

In the same way, an equation as simple as

$$x^2 = 2$$

lacks a solution in the field of rational numbers. This was the problem faced by the Greeks in the ancient world. However, they had to accept such a monster because even the solution to the simple geometric problem of finding the diagonal of a square ran into this very beast.

The solution to that equation amplified the numeric field, expanding it to include so-called real numbers:

$$x^2 = 2 \iff x = \sqrt{2}.$$

At this point, the mathematical mind reacts by thinking that it is not that the equations do not have a solution because the numbers that express it do not exist, but that every equation has a solution. The problem arises when that solution belongs to an already known or assimilated numeric field. Let's take things a step further: the equation

$$x^2 = -1$$

has no solution.

But it does not have a solution because we take the x to be a real number, a finite or infinite, periodic or non-periodic decimal. The value of x that verifies the equation seems utterly monstrous:

$$x^2 = -1 \Leftrightarrow x = \sqrt{-1}.$$

In the middle of the 16th century, Gerolamo Cardano found the formula for resolving cubic equations. But in applying the equation $x^3 - 15x - 4 = 0$ he found himself faced with another surprise. It is easy to show that $x = 4$ is one solution. However, the solved equation looked very different:

$$x = \left(2 + \sqrt{-121}\right)^{\frac{1}{3}} - \left(-2 + \sqrt{-121}\right)^{\frac{1}{3}}.$$

We have come across the monsters again. What is the meaning here of the square root of a negative number? How can such numbers be related in something as straightforward as $x = 4$? If we have to accept the square roots of negative numbers as numbers, what meaning can we ascribe to them?

In the early 19th century the roots of negative numbers found a full meaning and significance by being expressed in Cartesian coordinates on a third axis (z in addition to the x and y of the real number field), and from there became an integral part of the so-called complex numbers. Represented by the symbol $\mathbb{C}$, the set of complex numbers extends the field of numbers. A complex number is a number composed of two parts, one real and another imaginary. The imaginary part is the product of a real number multiplied by i, the root of minus one, also called 'imaginary root of unity'. Let's take two complex numbers, a and b:

$$\begin{aligned} i &= \sqrt{-1} \\ a &= 2 + 3i \\ b &= \frac{1}{2} - i\sqrt{5}. \end{aligned}$$

A number such as $a = 2 + 3i$ is represented on the Cartesian plane taking two units on the abscissa (x) axis and three on the ordinate (y) axis. The resulting point has the coordinates (2, 3). But it is not only a point that we have represented because, unlike the points and vectors on the plane, with complex numbers all known algebraic

operations (sums, subtractions, quotients, products, powers, etc.) can be carried out. Also, by calculating with complex numbers things work in a similar way to how they do with real numbers. And to top it off, it is a complete system, as any equation proposed in the complex field has a solution in it, something that does not apply to the other number systems.

Since complex numbers found their place on the plane they have been crucial for solving problems which were irresolvable with the real numbers system.

Symbiosis of algebra with geometry

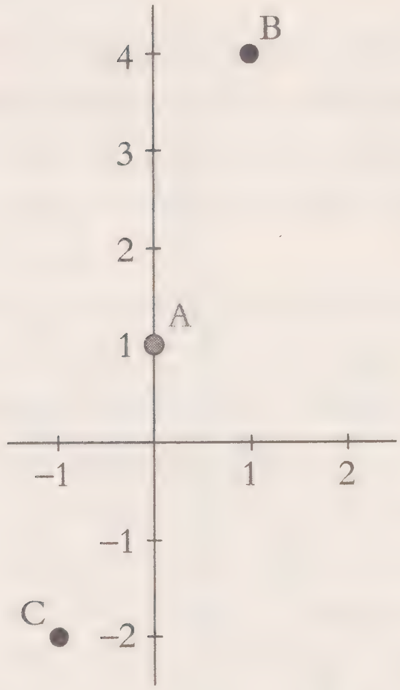
If the above was possible it is thanks to another great creation, the symbiosis of algebra with geometry by means of the analytical geometry progressed by Descartes and Fermat. In ancient times there had already been some attempts that could be interpreted as the beginnings of systems for the geometrical representation of formulae. But it was Descartes who joined the hands of algebra and geometry and they never let go again.

The world of algebra is about formulae and equations. Geometry focuses on shapes and space. In analytical geometry the two become one. For each shape there is a formula that describes it; for each formula there is a set of points on the plane with which to verify it. This lends the equations a visual and spatial feeling.

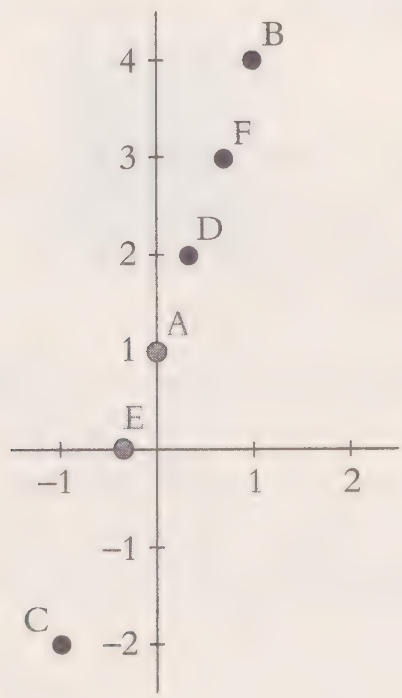
The advantage of analytical geometry resides in the fact that it localises its solutions and equations and places them on the mathematical map using a system of coordinates. Its value does not come from the ability to develop proofs, as some geometric theories are expressed more elegantly, briefly and clearly by Euclidean demonstration than by analytical demonstration.

The equation $3x - y + 1 = 0$ is an algebraic element whose original meaning and significance refers to two numbers, x and y, that satisfy that equation. There are various pairs that fit the equation: $x = 0$, $y = 1$; $x = 1$, $y = 4$; $x = -1$; $y = -2$.

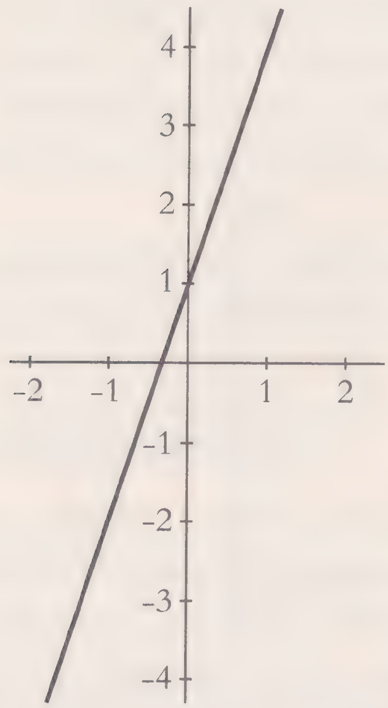
Proportional analytical geometry provides the situation with new meaning, and this is achieved by quantifying space. If the space is the two-dimensional plane, two straight lines are taken on which the real numbers are represented, one corresponding to each dimension. For convenience they are drawn perpendicular, although this is not essential. Creation continues to associate the values of variable x with those on one axis, and those of variable y with those of the other. Thus, we mark points A, B and C on the plane, corresponding to the three pairs of solutions given above:



Adding more pairs that fit the equation the shape taken on by the set of solutions can be seen:

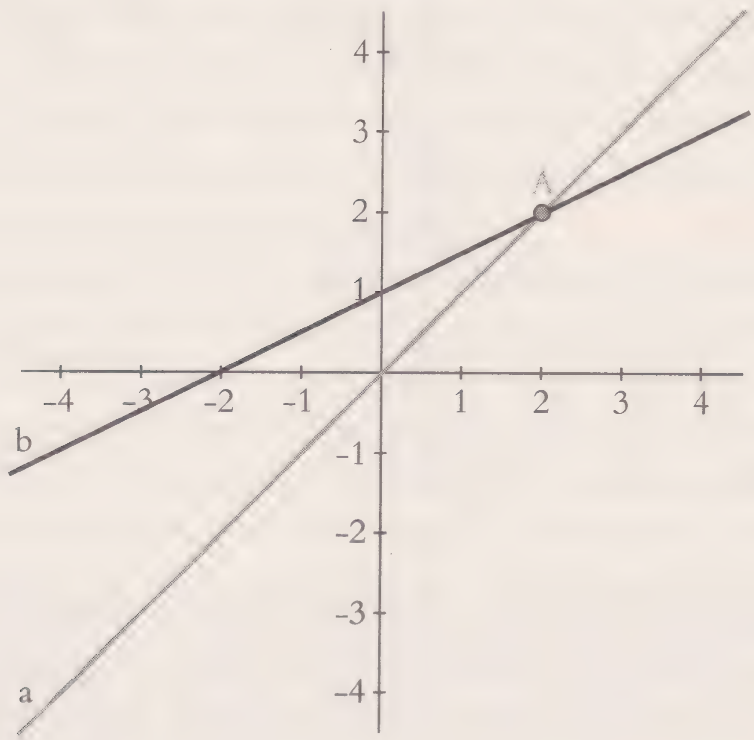


All we need to do is isolate one of the two variables to see that for each of their values there are others of the second variable that satisfy the equation. The infinity of possibilities of one implies the infinity of possibilities of the other. The result is that the algebraic equation $3x - y + 1 = 0$ corresponds to a straight line on the plane:



Consequently, a system of two equations with two unknown values becomes a geometric problem, finding the point of intersection of two straight lines.

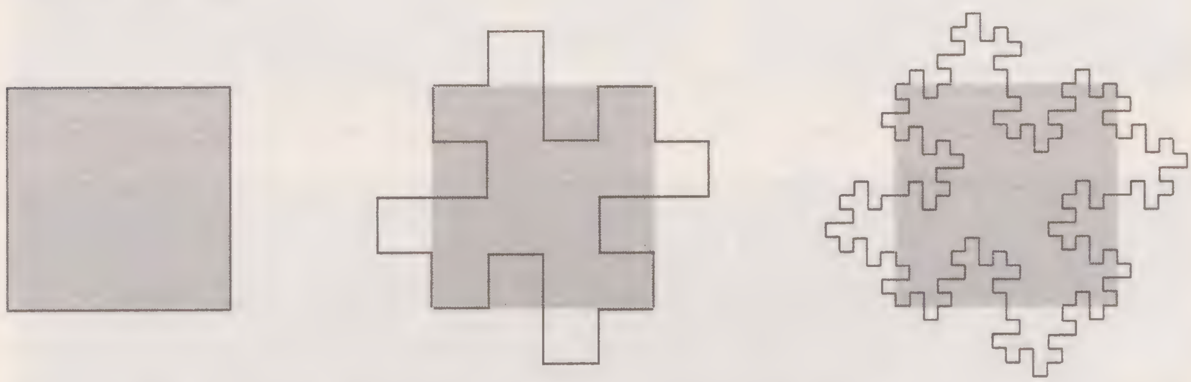
$$\left.\begin{array}{l} -x+2y-2=0 \\ -x+y=0 \end{array}\right\}$$



New technology, new curves

Mathematical creation has been highly influenced by the technological development of recent decades. Computers allow seemingly endless tasks that would take people centuries of work to be carried out in the matter of seconds. This, together with the ever-more powerful capacity of imaging software, has made the computer a test bed and mathematical microscope.

Thanks to this technology we have discovered fractals and invisible curves, and others that were barely imaginable fifty years ago. It is not that fractals were unknown, but their interest, the possibility of seeing them and using them have been expanding along with technological development. The origins of fractals can be found in the 'von Koch curve', also called 'Koch's snowflake'. Unlike traditional curves constructed by assigning values to a function, Koch's snowflake is created by a recursive algorithm. We start with a square, a triangle or any other figure which is replaced by a polygonal line. The process is then repeated applying the same change to each segment of the polygonal line to create a figure with an increasingly irregular profile.



The first in-depth study of the fractals was made by Benoît Mandelbrot, a Frenchman of Polish origin, in the 1980s. One of the key ideas of the fractal concept is that of 'orbit of a given point'. For all functions, such as, for example, $f(x)=x^2$, we can study the orbit of a given point or the series of results which are obtained by recurrently substituting the variable with that point:

$$
\begin{aligned}
&f(0.5)=0.5^2=0.25\\
&f(0.25)=0.25^2=0.0625\\
&f(0.0625)=0.0625^2=0.0039\\
&\Rightarrow\ \text{Orbit of } 0.5=\{0.5;0.25;0.0625;0.0039;\ldots\}\rightarrow 0.
\end{aligned}
$$

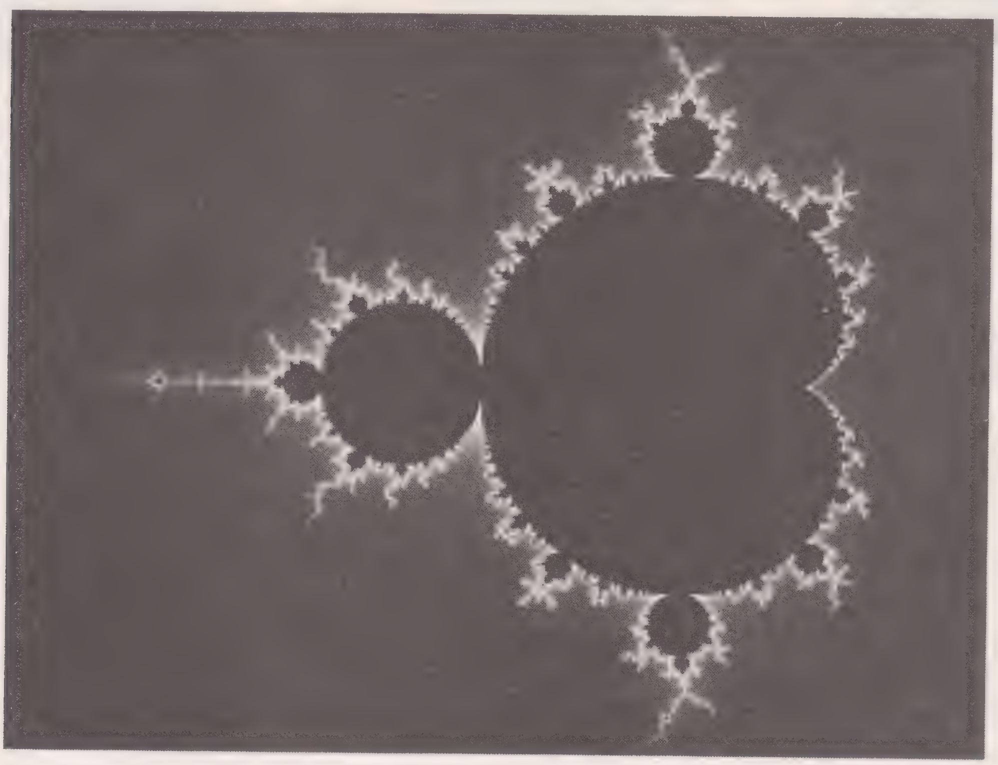
The orbit of point $x = 0.5$ is formed by a decreasing and unlimited succession which tends towards 0. There are other fixed orbits such as those of $x = 0$ and $x = 1$. And there are those that escape towards infinity, such as $x = 2$:

$$x = 2$$
$$f(2) = 2^2 = 4$$
$$f(4) = 4^2 = 16$$
$$f(16) = 16^2 = 256$$
$$\Rightarrow \text{Orbit of } 2 = \{2,\ 4,\ 16,\ 256,\ \ldots\} \rightarrow \infty.$$

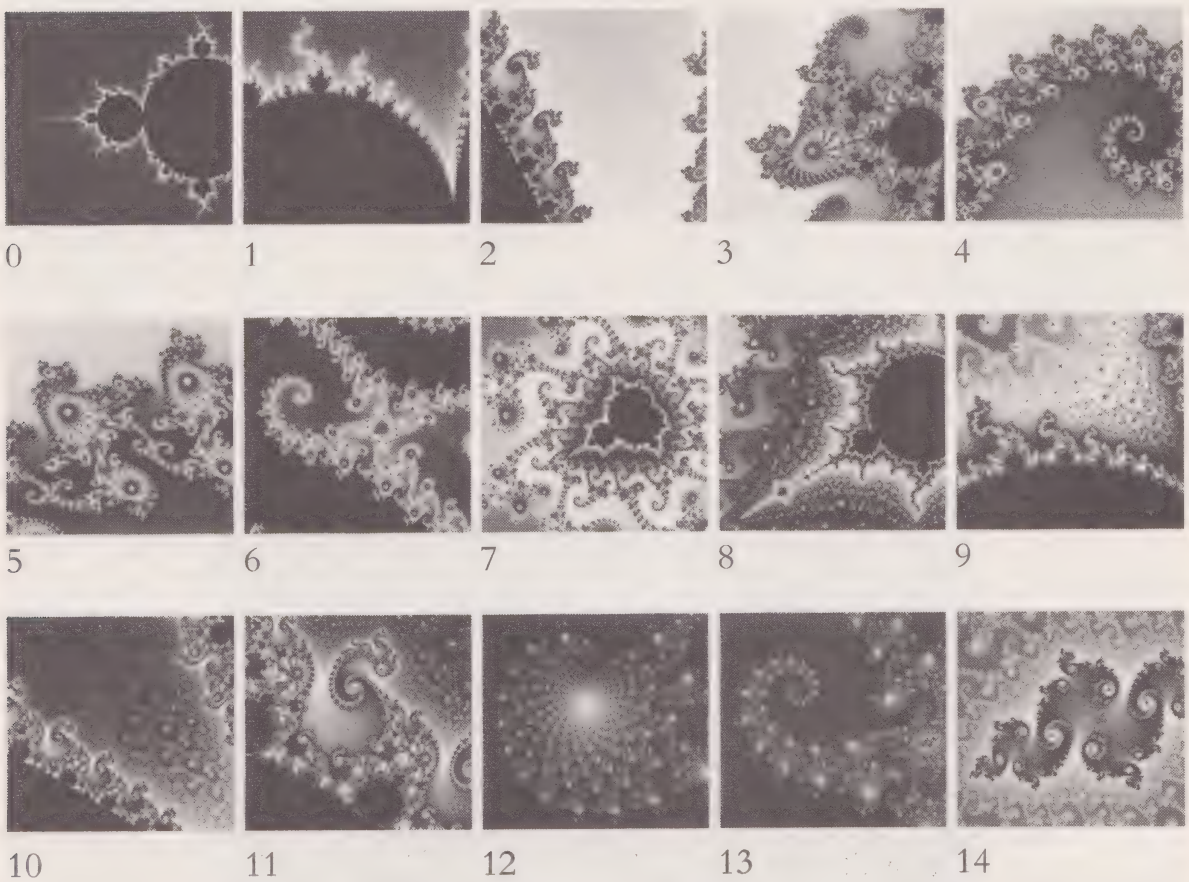
The computer allowed us to see what would happen with a function similar to this, but in the field of complex numbers:

$$f(z) = z^2 + c \text{ , being } z, c \in \mathbb{C}.$$

The result was unexpected both from a mathematical and aesthetic point of view, as the shapes adopted by the set of points which did not escape towards infinity according to the values of c were as varied as they were surprising. These points form the so-called 'Julia set'. The complex values c for which the Julia set is a connected set, that is, made of one piece with no fragments, form the Mandelbrot set M, whose strange appearance is as follows:



The Mandelbrot set was not seen until 1980. There is no object in mathematics as complex as this. Apart from its fractal character, by which each extension of one of its parts resembles the whole, it has limitless variations. Effectively, if we zoom in on any zone of the set and observe the new image resulting from the increased scale, we will see the same figures reproduced over and over again.



The M set combines self-similarity and change in an infinite spiral; a genuine example of creativity.

From a topological point of view, a fractal curve is different to traditional curves. Specifically, the main difference resides in its infinite self-similarity. The amplification of a traditional curve around a point results in a segment, while, on the other hand, any amplification of a fragment of a fractal has the same appearance as the original fragment. This means that the dimension of a fractal object is not an integer between one and three, as with 'traditional' objects. The dimension of the Von Koch curve, for example, is 1.26186... In fact, and despite the fact that computers allow us to see various stages of fractal objects, we will never be able to see the complete process,

as it has infinite levels. The fractal reality impedes us from seeing the profiles crisply. As soon as we try to focus on them we realise that they have changed, and they are no longer as they were before.

EDIBLE FRACTAL

Fractal shapes are so common in the real world that we can talk about the fractal geometry of nature. However, they do not often exceed four levels of self-resemblance. This is the case with the branching of vegetables, nerves and rivers. The fractal dimension is a characteristic that is useful for diagnosing bone diseases and to characterise encephalograms. The cauliflower shown here is actually a hybrid which was identified for the first time in Italy in the 16th century. Its structure constitutes an extraordinary example of natural fractal geometry. The whole piece (level 1) is formed by reduced copies of itself aligned in spirals. In turn, each of them is composed of other reduced copies also arranged in a spiral (level 3). The same is repeated again for another, higher level (4).

Chapter 3

New Questions for Common Situations

In the previous chapter we covered history's greatest mathematical creations. Nowadays the task of developing mathematical knowledge is the responsibility of professionals, but they are not the only ones. Creating mathematics is not just producing the great theorems that go down in history, but also proposing problems, offering mathematical explanations for phenomena, developing practical methods for translating mathematics into reality, using technology to create mathematics and mathematical solutions and, above all, realising that the mathematical answer to a question is necessary and satisfying. Many people are capable of creating mathematics. It may be that their results and the conclusions they will reach are not new to the professionals, but, essentially, their work is the same.

In this chapter we present mathematical creations related to a diverse range of fields. Most of them are unrelated to the academic world, and are the result of proposing new questions, looking for different solutions, establishing new meanings to long-established concepts and contributing mathematical ideas to new contexts. We live through creation. If we ask ourselves mathematical questions, we will create mathematically.

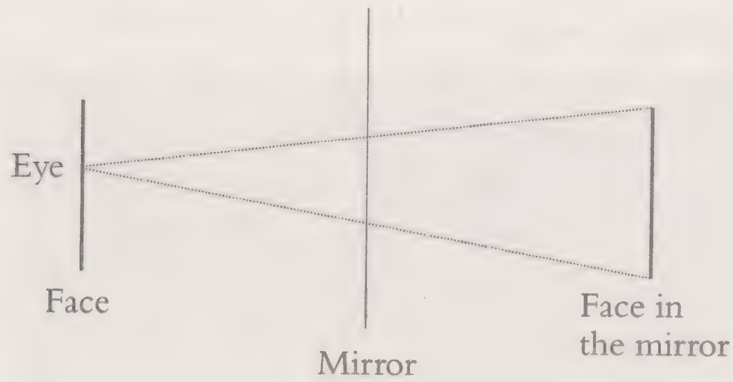
Where do we start? How can we be creative in mathematics? The first inroads are found in everyday life. We are going to study some phenomena which we have all experienced once in a while but few of us have approached mathematically. From here, we will remove ourselves from reality to end up in purely mathematical realms.

The first thing to do is to put ourselves in the situation and contemplate the phenomenon (object, person, process, experience) that will capture our attention with mathematical perspective. This means proposing objective questions, the answers of which will not depend on our taste, convenience and appetite, but which should be argued on spatial or relative qualities which are in some way quantifiable.

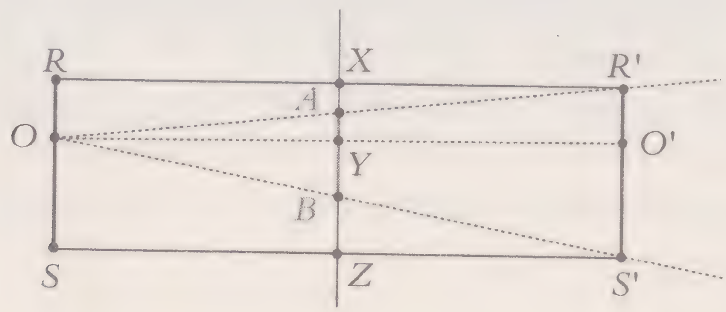
Every day in front of the mirror

Immediately after waking up, the first thing we normally do is go to the bathroom to wash ourselves. We wash our face, put our make up on, comb our hair, etc. in front of the mirror. We do this every day of our lives. And what do we ask of the mirror? Not much. Just that it reflects our complete face. After breakfast, and just before closing the front door, maybe we have a look at the full-length mirror in the hallway to see how we look. The requirements for this one? That it shows us our complete body.

How many times have we done this and how many times have we asked ourselves for the minimum dimensions a complete bathroom or full-length mirror must have in order to fulfil its function? Never? Let's imagine that we are in front of a mirror in which we can see our entire face. What is the minimum height it should have? We start by drawing a geometric diagram of the situation, reducing the elements involved (points and segments) to their minimum expression.



The diagram already tells us the minimum dimension of the mirror. We need to find out what relation this has with the face reflected in it. To do so we reduce the diagram further, drawing the additional lines that are fundamental to the problem and adding notations for the main points in the arrangement.



Given that reflection $R'S'$ is symmetrical to its original RS, and the virtual image is created at the same distance but on the other side of the mirror, we have $RX = XR'$. And also, $RX = RR'/2$.

Also triangles OAY and $OR'O'$ are similar due to their parallel sides. The same applies to OYB and $OO'S'$. As $RX = RR'/2$, the ratio of magnification between these triangles is 2, therefore $AY = R'O'/2 = RO/2$ and also $YB = O'S'/2 = OS/2$.

In other words, $AB = AY + YB = RO/2 + OS/2 = (RO + OS)/2 = RS/2$. Therefore, the mirror must be equal to half the height of the face, at least. The height at which it needs to be hung is $BZ = YZ/2$, in other words, half way between the eyes and the chin. Similarly, the height of a full-body mirror should be half the height our eyes, and would have to be hung at half the height of our eyes too.

Contemplating the horizon

If we leave the house, perhaps we are lucky enough to be able to go down to the beach and contemplate the horizon. Who has asked themselves mathematical questions while looking at the horizon? Generally, admiring a sunrise or sunset inspires questions about the past and the future, about how beautiful nature is, about the fishermen returning to port with the day's catch, about what is hidden by the line behind which, more quickly than expected, the flaming orb that illuminates our lives is disappearing…

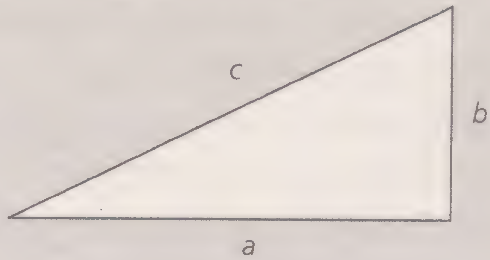
But if we adopt a mathematical point of view, questions and observations which come to our minds are of another type. If I crouch down, the horizon draws nearer; if I stand up, it moves away. If I turn round 360° on my feet I can see that the horizon is circular; that is why sailors and fishermen sensed that the world was spherical, because when they moved away from the shore to the point where they could not see land, they saw a circular horizon surrounding them. What distance separates me from the horizon? What is its radius? What distance separates me from another vessel which is moving across it?

This is the first step necessary for mathematical creation: asking ourselves things about what we experience, see and do, the answers to which transcend the personal and are aimed at the objective. It is, in brief, the scientific perspective.

From the mathematical point of view, the Earth is a huge sphere with a radius of about 6,370 km. The visible horizon for a human is the furthest point of the planet's surface reachable by sight. Supposing that light moves in a straight line, we

PYTHAGORAS' THEOREM

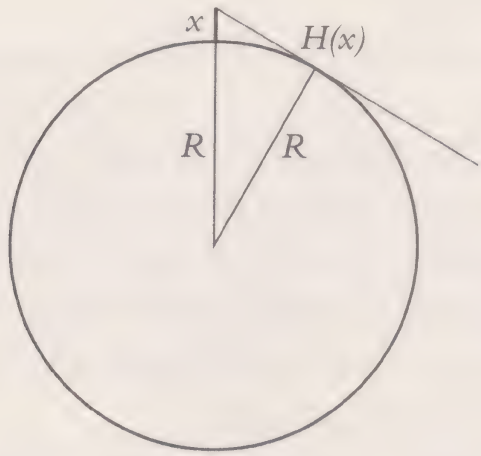
Triangle *abc* is a right-angled triangle $\Leftrightarrow c^2 = a^2 + b^2$.



This is the best-known theorem in mathematics, quoted and demonstrated daily in schools across the world. But the usual demonstration is not the Euclidean one (Proposition 47 of 'Book I' of *Elements*), but one based on the decomposition of a square as if it were a jigsaw puzzle.

The theorem refers to areas, although it has always been used to calculate lengths. The necessary implication ($\Rightarrow$) is always demonstrated; never the sufficient one ($\Leftarrow$), although it is applied sometimes. In *Elements*, it follows the one above (Proposition 48).

create a geometric model of the real situation. In it, the horizon is determined by the tangent to a circle.



Where *H(x)* is the distance that separates us from the visible horizon, *x*, the height of our eyes, and *R*, the radius of the Earth, we have a right-angled triangle to which we can apply Pythagoras' theorem:

$$H(x)=\sqrt{(R+x)^2-R^2}=\sqrt{R^2+2Rx+x^2-R^2}=\sqrt{2Rx+x^2}.$$

The horizon contemplated by a person whose eye height is $x=1.7$ m and who is standing up on the shore, is at a distance of $H=4{,}653.8$ m ($R=6.370$ km).

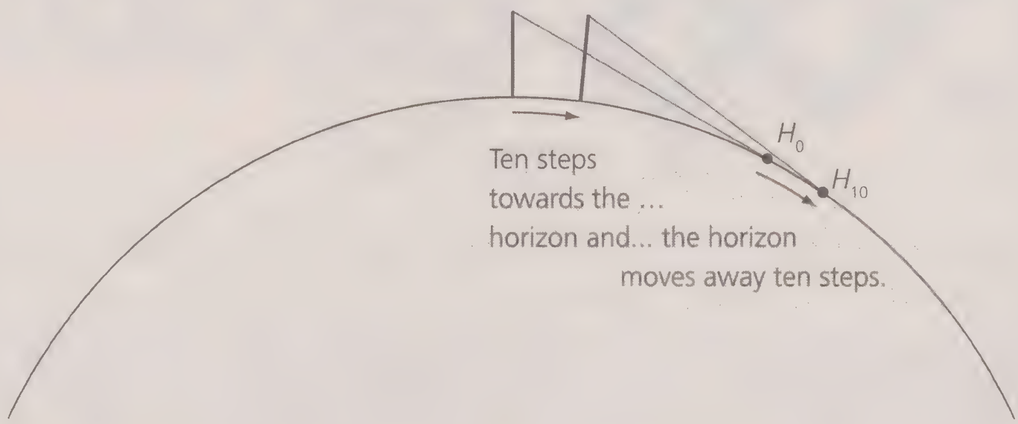
Where is the mathematical creation here? Have we created new mathematics? We have applied a powerful theorem to obtain an unknown formula that did not exist previously. Here we have the first mathematical creation related to this phenomenon. But it is not the only thing that is important. The main creation is in the question: how many times have we contemplated the horizon without ever asking ourselves how far away it is? Secondly there is the creation of a geometric model that allows the application of a mathematical result. It is the mathematical focus that makes us think of the Earth as a sphere; the line of sight, as a straight line, and our body, as a very short lengthening of the radius of that sphere. Also, we formulate a reduction to the bi-dimensional plane of a three-dimensional reality, as we reduce the spherical planet to a circle.

ON THE HORIZON

In his work *Cycles*, Spanish writer F.M. sees a good spot to place utopia on the horizon:

> "What is utopia for? It is on the horizon. If I move ten steps closer, the horizon moves ten steps away. That is what it is for, for learning to walk."

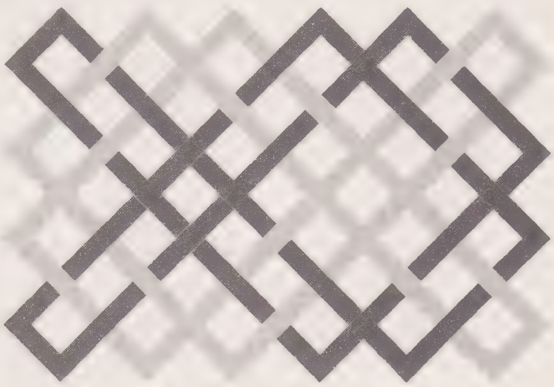
The statement is true enough according to how steps are counted on a sphere:



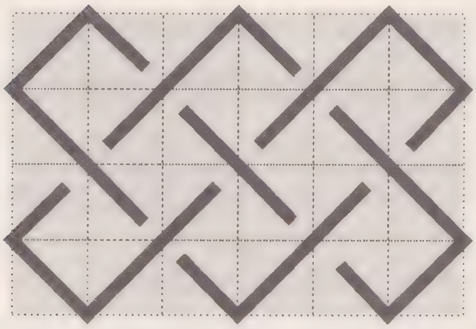
Cyclic knots

In his work *The Sense of Order*, Austrian art historian Ernst Gombrich explains how he draws Celtic knots. These are characterised as covering a grid in such a way that, after passing through all the points marked on every side, they return to the point of origin. In other words, there is an infinite knot with no end because it traces a route that ends exactly where it starts:

Knots of this type are not always infinite or cyclic:



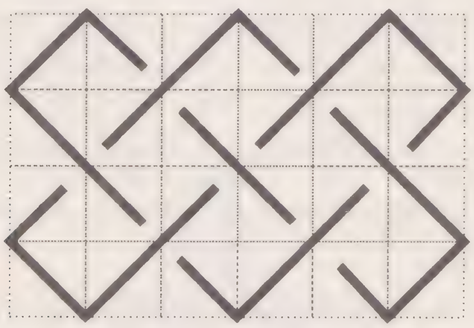
A question arises: why are some knots infinite and others not? Before suggesting possible answers it would be good to know how they are constructed. The construction is based on a grid with a series of points to which the knot is hooked, as if the knot were sewn into the background grid through those points.

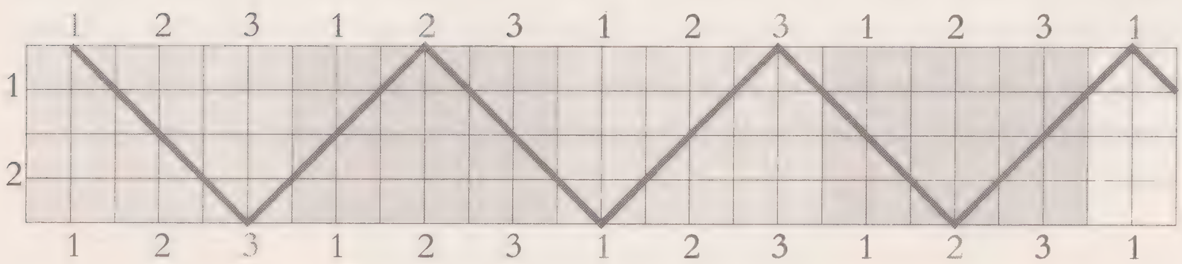


This allows the knots to be characterised based on the number of vertices on each of the sides of the grid which is used as a reference. The first example would be a 3×2 knot; the second, 3×3, and the last, 6×4. The 3×2 is sewn onto a 6×4 grid, the sides of which use horizontal vertices 1-3-5 and vertical vertices 1-3. The 6×4 grid is interpreted as $(1+2\cdot2+1) \times (1+2+1)$. The same is true for the rest. The 3×3 knot is on a grid of $6\times6=(1+2\cdot2+1)\times(1+2\cdot2+1)$, and the 6×4, on one of $12\times8=(1+2\cdot5+1)\times(1+2\cdot3+1)$.

We would say that the result depends on the number of vertices of the knot on each side of the grid. The 3×2 knot is infinite because it is made of one unique loop. The 3×3 knot is not infinite – it has three loops. The 6×4 knot is not infinite either, as it has two loops.

Where is the key? The thread moves left and right and up and down. If, instead of sticking to the rectangle, we had the chance to advance, vertically or horizontally, we would be able to read between the lines of the issue. Let's do this for a (3,2) knot:

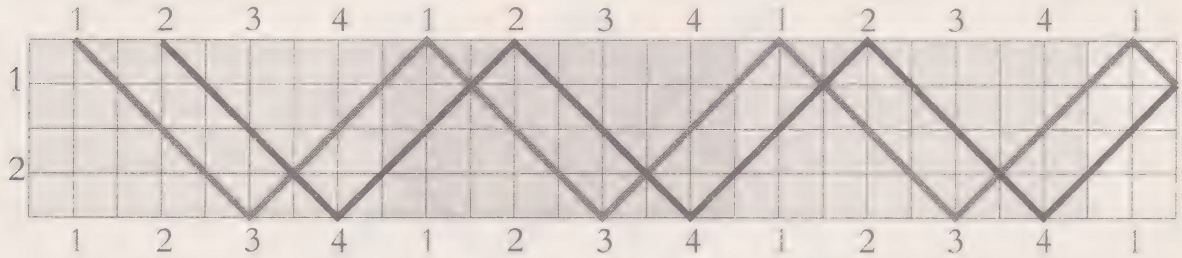




Starting from 1, and in steps of 2 units to the right, we arrive at 3, then 2 and, finally, at 1 again. A numeric cycle is created which is repeated *ad infinitum*:

$$[\,1, 3, 2\,] = 1, 3, 2, 1, 3, 2, 1, 3, 2, 1, \ldots$$

If the network is (4,2), two cycles are necessary:



$$[\,1, 3\,] = 1, 3, 1, 3, 1, 3, 1, \ldots$$
$$[\,2, 4\,] = 2, 4, 2, 4, 2, 4, 2, \ldots$$

In the first case, we skip in pairs over three. The cycle is completed after 6 steps, when we return to the point 1 where we started. Then we have passed through all the numbers 1, 2, and 3. In the second case we need two cycles to cover all the figures:

Grid										Cycles
(3,2)	1	2	3	1	2	3	1	2	3	
	*		*		*		*			[1,3,2]
(4,2)	1	2	3	4	1	2	3	4	1	
	*		*		*		*		*	[1,3]
		*		*		*		*		[2,4]

Why? Because here 4 can be divided by 2. If we start at 1 and jump in pairs, we pass through the 1 and the 3, never through the 2 and the 4. To do that we need to start a new route from 2. In the first case the cycle was completed after $6 = \mathrm{lcm}(3,2)$ [lcm means lowest common multiple] stages in a single cycle, as $\mathrm{lcd}(3,2) = 1$ [lowest common denominator]. The same applies to the example with the 6×4 network, with $\mathrm{lcd}(6,4) = 2$ cycles and the (3,3) network, which contained $3 = \mathrm{lcd}(3,3)$.

In conclusion:

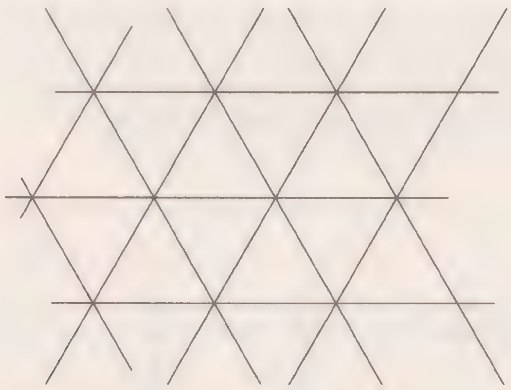
Theorem: *In a network of* (m,n) *vertices, the number of cycles is* lcd(m,n).

Corollary 1: *If* m *and* n *are coprime, the* (m,n) *network has a unique infinite cycle.*

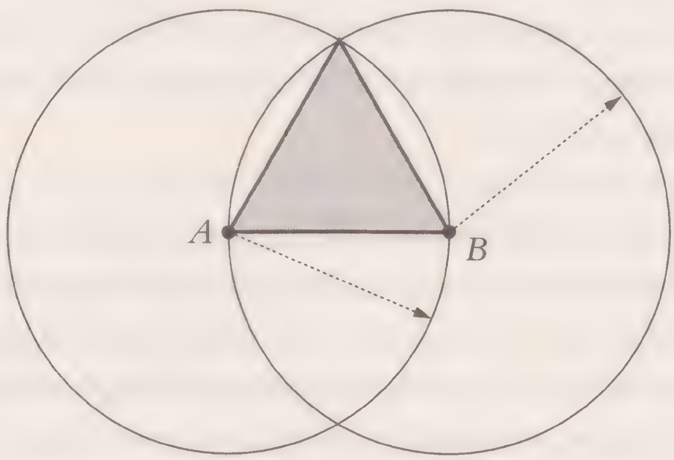
Corollary 2: *In a network of* (m, n) *vertices, the number of stages is* $2\cdot(m+n)$.

One on gardening: the equilateral triangle as a particular case of the isosceles triangle

Sometimes academic mathematical methods are not sufficiently practical to be applied to real situations. Planting in a staggered formation involves planting a tree at the vertices of a triangular and equilateral network which guarantees an equal distance between the planted trees.



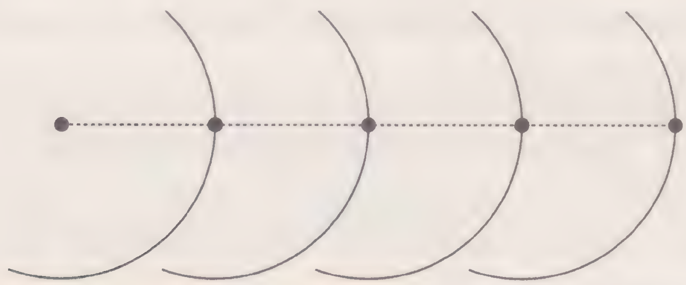
If a mathematician is to draw this network, the norm is for him to think of a formal way to trace equilateral triangles. He will almost certainly come up with a Euclidean solution, such as the method for the construction of the equilateral triangle of Proposition 1 of 'Book I' of *Elements*.



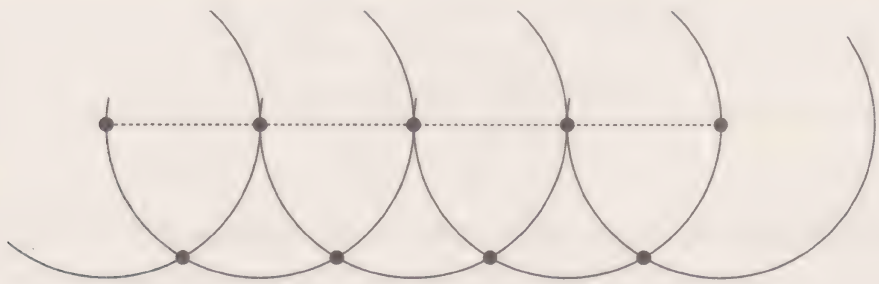
Proposition No. 1 of Euclid's Elements*: construction of the equilateral triangle on a given line segment with ends A and B.*

To carry it out he will change the two-armed compass for a piece of string of a length equal to the desired equilateral triangle. The mathematician will pass over the ground drawing circular arcs. Their intersection will set out the precise distribution of trees.

First, he will indicate a series of equidistant, aligned points:



Then, with the centre of each of them, he will trace circular arcs that are big enough to cut the vertices of the equilateral triangles:



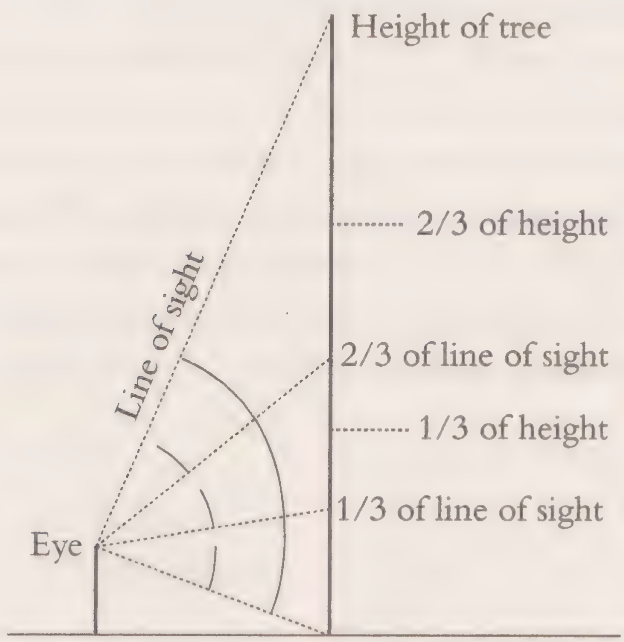
And so on, reapplying the same procedure to the last row and completing the network. That is what the mathematician would do, but that is not what a professional gardener does. According to Gil-Albert (1999, p. 84), the horticulturist creates the same network in a different way:

> "In the case of staggered plantations... To complete the pattern, one worker holds the end of a tape measure at the first pole of the main row. Another worker reels out the tape to measure out the planting distance, let's say 5 m. He places the next pole there, and then reels out another 5 m of tape. A third worker picks up the tape at the 5 m mark and walks backwards with it. When the tape is pulled tight, this third man will be 5 m from both his colleagues and so places a third pole at the vertex of an equilateral triangle."

The theoretical solution to a problem tends not to be the best practical solution. The professional mathematician's method is not the one applied in practice. From a mathematical perspective, the real-world method does not matter. However, the creation of the practical solution to the problem was not his, but that of the horticulturist. And creating a practical solution to a mathematical problem is a mathematical creation in itself.

Assisting foresters: a third of what we see is not a third of what we are looking at

Most tree pruning is carried out on lower branches on its bottom third. Given that the forester needs to estimate that lower third, a mathematical problem is created. Is the third of what we see a third of what we are looking at? Generally, the answer is no:

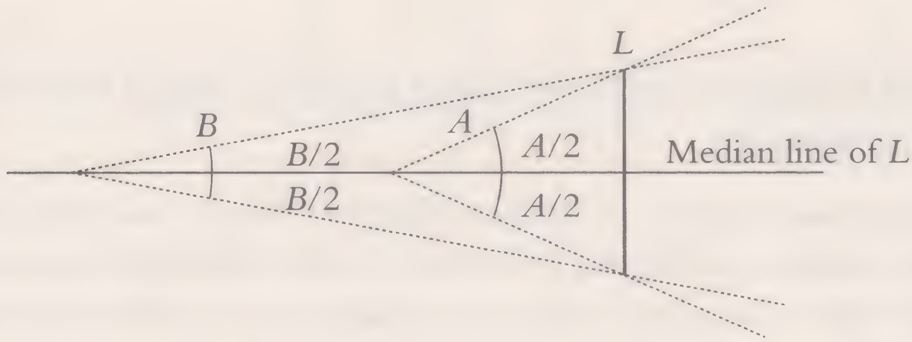


Only if we are looking at a circular arc from its centre will we find a coincidence between a third of what we perceive visually and a third of the object that we are looking at. So, what should the forester do? How can he visually determine a third of the tree's height?

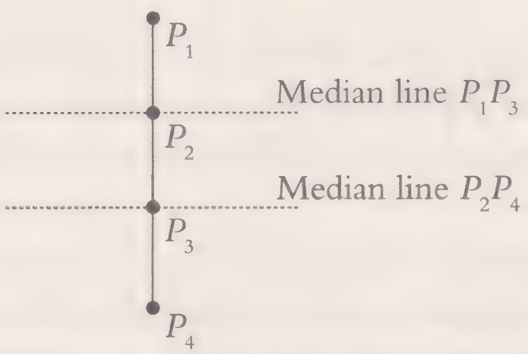
Normally we do not know the height of the tree because it is too high to measure easily. If A_1 is the angle of vision with which the whole tree is perceived (a theodolite is required); a, the height from which the observer is looking and d, the distance to the foot of the tree, then angle A_3 which determines the lower third of its height is:

$$A_3 = \arctan\left(\frac{d \cdot \tan(A_1) - 2a}{3d}\right).$$

What is the key to the problem? Our perception sees angles. Only from a point on its perpendicular bisector will a half of the line that we are looking at correspond to a half of the line that we see:



The same is not true of a third. If it were possible, a point X would exist on the plane from which the thirds P_1P_2, P_2P_3 and P_3P_4 of the segment P_1P_4 would be seen from the same angle (see the diagram below). So, because from X the two halves of P_1P_3 would be seen from the same angle, X should be on the perpendicular bisector of P_1P_3 (straight line through P_2 perpendicular to P_1P_3). The same would apply to the perpendicular bisector of P_2P_4 (straight line through P_3 perpendicular to P_2P_4). Then X would be on both perpendicular bisectors, which would be parallel as they would be perpendicular to the same segment P_1P_4. Sadly that's impossible:



Apart from the case of observing a circular arc from its centre, the third which we see is not the third we are looking for.

Warning from accountants: the rounded sum is not the sum of its rounded parts

Rounding establishes that if the last decimal figure is less than 5, that decimal becomes a 0, and if it is greater than five, the previous one increases its value:

$$2.34 \approx 2.3.$$
$$2.37 \approx 2.4.$$

Errors in approximation as small as tenths, hundredths and thousandths can be very significant when dealing with them in large quantities. An error of one pence in a current account could be considered insignificant, but when it is multiplied by 300 million people or accounts it becomes £3 million! Accounting is a field of mathematics where errors cannot be overlooked. In calculating balances, cents can dance, that is, they can fall on one side or the other. This happens because, for example:

$$0.3+0.4=0.7 \approx 1.$$
$$\left.\begin{array}{l} 0.3 \approx 0 \\ 0.4 \approx 0 \end{array}\right\} 0+0=0.$$

We have a theorem: *The rounded sum is not the sum of its rounded parts.*

This statement is outlined graphically in the tables below:

Rounded sum										
	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	1	1	1	1	1
1	0	0	0	0	1	1	1	1	1	1
2	0	0	0	1	1	1	1	1	1	1
3	0	0	1	1	1	1	1	1	1	1
4	0	0	1	1	1	1	1	1	1	1
5	0	0	1	1	1	1	1	1	1	1
6	0	0	1	1	1	1	1	1	1	2
7	1	1	1	1	1	1	1	1	2	2
8	1	1	1	1	1	1	1	2	2	2
9	1	1	1	1	1	1	2	2	2	2

Sum of the rounded parts										
	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	1	1	1	1	1
1	0	0	0	0	0	1	1	1	1	1
2	0	0	0	0	0	1	1	1	1	1
3	0	0	0	0	0	1	1	1	1	1
4	0	0	0	0	0	1	1	1	1	1
5	1	1	1	1	1	2	2	2	2	2
6	1	1	1	1	1	2	2	2	2	2
7	1	1	1	1	1	2	2	2	2	2
8	1	1	1	1	1	2	2	2	2	2
9	1	1	1	1	1	2	2	2	2	2

Note the frequency with which the results 0, 1 and 2 appear in the tables:

	0	1	2
Rounded sum	15%	70%	15%
Sum of the rounded parts	25%	50%	25%

Why not correct those differences and determine a system for rounding which provides a more balanced distribution of 0s, 1s and 2s? For example, each of the results could be set at around 33.3%, which would mean distributing zeroes, ones and twos in the 100 cells in a symmetrical fashion. This is shown in the following table, in which figures 0, 1 and 2 appear 33, 34 and 33 times, respectively:

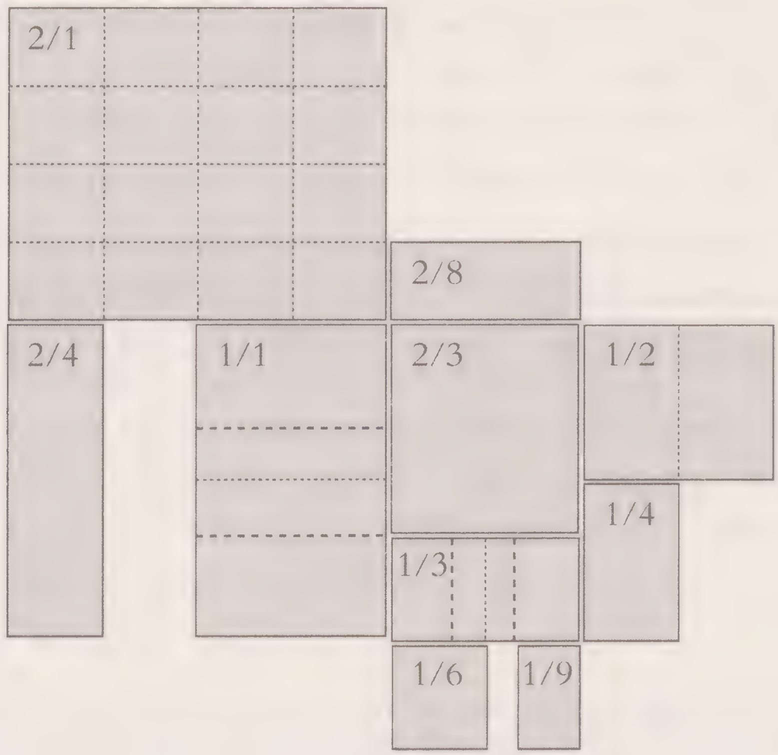
Rounded (Sum) = Sum (Rounded parts)										
	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	1	1	1
1	0	0	0	0	0	0	1	1	1	1
2	0	0	0	0	0	0	1	1	1	1
3	0	0	0	0	0	1	1	1	2	2
4	0	0	0	0	1	1	1	2	2	2
5	0	0	0	1	1	1	2	2	2	2
6	0	0	1	1	1	2	2	2	2	2
7	1	1	1	1	2	2	2	2	2	2
8	1	1	1	1	2	2	2	2	2	2
9	1	1	1	2	2	2	2	2	2	2

Fridge Tetris

European cold storage rooms are optimised thanks to a standard tray format called Gastronorm EN 631. All the containers designed according to this standard are rectangular and identified by means of a numeric code which shows the ratio of its dimensions compared to a master tray size from which the rest are derived. These codes are:

2/1 2/3 2/4 2/8 and 1/1 1/2 1/3 1/4 1/6 1/9.

The master tray is the 1/1, which has dimensions of 530 mm × 265 mm. The remaining trays are made from it by duplication and division in the way shown in the diagram.



This means that the name of each tray expresses its relationship with the size of the 1/1 tray:

2/1 = double 1/1.
2/4 = a quarter of 2/1 = half of 1/1.
2/8 = an eighth 2/1 = a quarter of 1/1.
2/3 = two thirds of 1/1.
1/2 = half of 1/1.
1/3 = a third of 1/1.
1/4 = a quarter of 1/1.
1/6 = half of 1/3.
1/9 = a third of 1/3.

But these equalities are not a mere code. They are also true in terms of the calculations they encompass:

$$\frac{2}{1} = 2 \cdot \frac{1}{1} = 2.$$
$$\frac{2}{4} = \frac{1}{4} \cdot \frac{2}{1} = \frac{1}{2} \cdot \frac{1}{1} = \frac{1}{2}.$$
$$\frac{2}{8} = \frac{1}{8} \cdot \frac{2}{1} = \frac{1}{4} \cdot \frac{1}{1} = \frac{1}{4}.$$
$$\frac{1}{9} = \frac{1}{3} \cdot \frac{1}{3}.$$

Therefore, the Gastronorm codes are, in fact, numerical fractions, and constitute a representation system that reflects the size relationship between the containers. In order to find out, for example, how many 1/6 trays equal the 2/3 tray, all we have to do is calculate the division:

$$\frac{\frac{2}{3}}{\frac{1}{6}} = \frac{2 \cdot 6}{1 \cdot 3} = \frac{12}{3} = 4.$$

The Gastronorm system becomes a chilly game ofTetris which allows a particular rectangular storage space to be optimised in advance, fitting trays together as if they were pieces of a puzzle.

An infinite book and a two-dimensional disk

In the world of literature there are several writers capable of capturing, defining and illustrating mathematical ideas with great clarity. This contributes to the comprehension of mathematical concepts and provides another perspective from which to conceive them. To illustrate this point we are going to comment, from a mathematical point of view, on two stories, one by the Argentinian, Jorge Luis Borges, and the other by the Italian, Italo Calvino.

Most of Borges' work is developed around paradoxical situations, which are so logical that they seem real. The desert as a maze without doors or passages, the maze of mazes. A library with a tremendously complex arrangement. It is in the description of these objects and ideas where the tale touches on mathematical ideas.

In *The Book of Sand*, the Argentinian writer talks about a book with infinite pages in which neither the first nor the last can be found. From the text we could say

that Borges is thinking about potential and countable (or numerable) infinities, as all the pages which he speaks of have natural numbering. There are so many that it is impossible to open it on exactly the same page as one already seen, a detail which demonstrates an essential distinction between the finite and the infinite.

> "I opened it at random...I noticed that one left-page bore the number (let us say) 40,514 and the next, and odd one, 999. I turned the page; it was numbered with eight digits. It also bore a small illustration, like the kind used in dictionaries: an anchor drawn with pen and ink, as if by a schoolboy's clumsy hand.
>
> It was at that point that the stranger spoke again.
>
> 'Look at it well. You will never see it again.'
>
> There was a threat in the words, but not in the voice.
>
> I took note of the page, and then closed the book. Immediately, I opened it again. In vain I searched for the image of the anchor, page after page.
>
> ...The man who owned it didn't know how to read... He told me his book was called the *Book of Sand* because neither sand nor this book has a beginning or an end."

If the book were finite, as many pages as it may have had (for example, N), the probability of opening it again on a certain page would be small, but positive. On the other hand, in an infinite book, that probability becomes zero:

$$P(1) = \frac{1}{N} > 0. \qquad\qquad P(1) = \frac{1}{\infty} = 0.$$

We could consider that the *Book of Sand* could be numbered with the natural numbers: 1, 2, 3, ... Such numbering would impede ending the volume with a last page, but it would start it with a first page. However, the book lacks both. When someone searches for them, others always appear, both at the beginning and the end of the volume:

> "He suggested I try to find the first page.
>
> I took the cover in my left hand and opened the book, my thumb and forefinger almost touching. It was impossible: several pages always lay between the cover and my hand. It was as though they grew from the very book.

'Now try to find the end.'

I failed there as well...

'This can't be.'

[...]

'It can't be, yet it is. The number of pages in this book is exactly infinite. No page is the first page; no page is the last. I don't know why they're numbered in this arbitrary way, but perhaps it's to convey the idea that the terms of an infinite series can be numbered any way whatsoever.'

[...]

'If space is infinite we are in any point of space. If time is infinite we are at any point of time.'"

This lack of a first page invalidates our model of natural numbers. Which set of numbers fits the property of not having a first element? Positive rational numbers, that is, fractions, finite or periodic decimals. As well as being infinite and numerable (they can be counted), they lack a first and last number. Effectively, what is the first positive decimal before zero? There isn't one. If there were one, let's call it A, we could divide it by 2 $A/2$, a number which is still positive, rational and still smaller than A:

$$0 < \frac{A}{2} < A.$$

So the first should be $A/2$. But it cannot be this either because $A/4$ would come first. And $A/8$, even earlier. So, given a rational number (a page of the *Book of Sand*), any quantity of rational numbers (pages of the book) can lie between it and the 0 (the cover of the book). So we can number the pages of the *Book of Sand* with rational numbers between 0 and 1. There will not be a first page or a last page.

What does Borges mean when he says that in infinite space and time we find ourselves at any of its points or instants? Maybe this means we lose our reference points with their limits or extremes. In finite space or time we can talk about halves, thirds, proportions and distances to the extremes. In the case of infinity these ideas no longer make sense.

Borges has in mind a clear idea of infinity and its relationship with the different spatial dimensions:

JORGE LUIS BORGES (1899–1986)

A prominent writers of the 20th century, Borges' work is difficult to classify and spans genres such as narrative, essay, poetry and fantasy. Borges' fantasy is not without logic. In his tales there are good, accessible illustrations of ideas in the field of science and mathematics, which he demonstrates in a way that is understandable to the general reader. This is the case in *The Library of Babel, Funes the Memorious, The Analytical Language of John Wilkins* and *The Garden of Forking Paths*. In this work there are characters who have glimpsed the results of quantum mechanics.

On the back of the Argentinian two pesos coin, which commemorated the hundred-year anniversary of the birth of the writer from Buenos Aires in 1999, there is a labyrinth, a recurring concept in his stories.

> "The line consists of an infinite number of points; the plane of an infinite number of lines; the volume of an infinite number of planes; the hypervolume, of an infinite number of volumes…"

And in another story, dimension and not cardinality is the protagonist. *The Disk* is a very short story in which a greedy, murdering lumberjack spends years looking for an extraordinary object that is lost by his victim as he is struck by the killer's axe. It is the disk of Odin, which only has one face:

> "The lumberjack meets a traveller, who says: 'I wander the paths of exile, but still I am king, for I have the disk. Do you want to see it?'

He opened his hand and showed me his bony palm. There was nothing in it. His hand was empty. It was only then that I realized he'd always kept it shut tight.

'You may touch it.'

I had my doubts, but I reached out and with my fingertips I touched his palm. I felt something cold, and I saw a quick gleam. His hand snapped shut. I said nothing. He continued as if he were speaking to a child:

'It is the disk of Odin. It has but one side. There is not another thing on Earth that has but one side. So long as I hold it in my hand I shall be king.'

'Is it gold?' I said.

'I know not. It is the disk of Odin and it has but one side.'"

A three-dimensional disk has three faces. Two of them are circular and the third is the band that joins them and which we can imagine unwrapped into a rectangle. Two-dimensional objects do not have a thickness. Borges' mathematical creation is in proving that the disk of Odin lacks thickness as it is missing a face. The lumberjack never stops searching because the disk must have fallen face-down, in other words, the invisible face is face-up.

The quarters of Dorothea

We can find mathematical references in various works by Italo Calvino: *Cosmicomics*, *The Cloven Viscount*, *The Invisible Cities.* The last one is not, and does not attempt to be, mathematical, although the text contains many expressions that can be related to mathematical ideas. It relates how Marco Polo describes imaginary cities to the emperor Kublai Khan. Each city has a woman's name. We have chosen, as inspiration for mathematical creation, a sentence from the city of Dorothea:

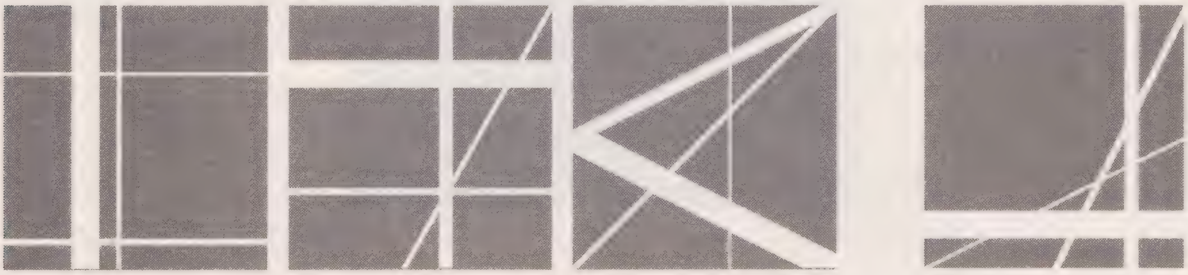
> "There are two ways of describing the city of Dorothea: you can say that four aluminium towers rise from its walls flanking seven gates with spring-operated drawbridges that span the moat whose water feeds four green canals which cross the city, dividing it into nine quarters, each with three hundred houses and seven hundred chimneys."

Calvino uses specific quantities to describe some architectural elements of the city of Dorothea: 4 towers, 7 gates, 4 green canals, 9 quarters, 300 houses and 700 chimneys. These figures invite us to make our first calculation on the city. It can be

said of Dorothea that there are $9 \cdot 300 = 2{,}700$ and $9 \cdot 700 = 6{,}300$ chimneys, which means that many houses have more than two chimneys.

But we are not going to base our creation on these calculations, but on a topological aspect such as the fact that the waters in the moat are fed by 'four green canals that cross the city, dividing it into nine quarters'.

Let's suppose that the canals are rectilinear. There are several ways of creating nine quarters with four canals, although up to eleven can be created, as shown in the following diagrams.



The question arises: what is the maximum number of quarters that can be created with a series of rectilinear canals or streets in a city? In other words, what is the maximum number of enclosures that can be created by n segments in a flat area?

To answer the question we note that the first street creates, at the most, two quarters and that the maximum quantity of these is obtained when the new straight street intersects all of the previous ones.

The first street creates a new quarter; the second, two more, the third three more and so on. Thus, the nth street added will create n new quarters. Therefore:

Street added	Old quarters	New quarters	Total quarters
0	1	0	1
1st	1	1	2
2nd	2	2	4
3rd	4	3	7
4th	7	4	11
5th	11	5	16

In other words, the maximum number of quarters $B(n)$ is obtained by adding n to the previous number of quarters $B(n-1)$:

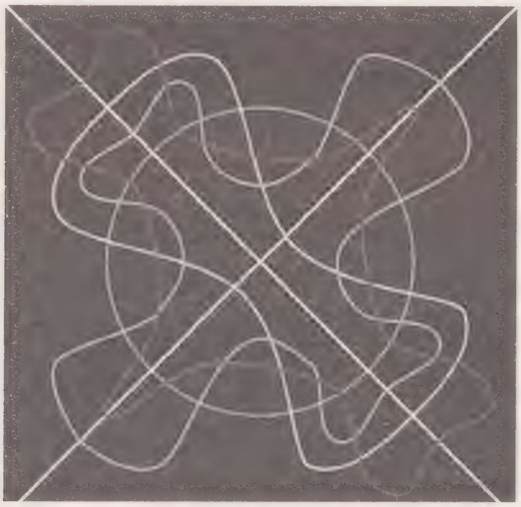
$$1.$$
$$1+1=2.$$
$$1+1+2=4.$$
$$1+1+2+3=7.$$
$$1+1+2+3+4=11.$$
$$1+1+2+3+4+5=16 \;\Rightarrow\; B(n)=1+\sum_{i=1}^{n} i=\frac{n^2+n+2}{2}.$$

See the geometric configuration taken on by a similar city:

The suggested curve is the so-called envelope of $B(n)$ for $n \rightarrow \infty$. It is the hyperbole of the equation:

$$x^2+y^2+2xy-4y=0.$$

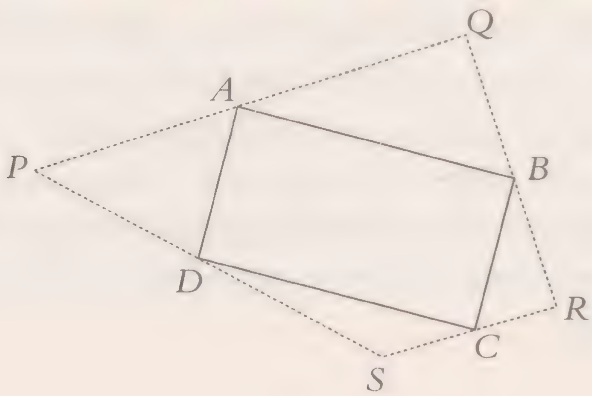
Supposing the streets are not necessarily straight: it can easily be seen that the maximum number of possible streets becomes $B(n) = 2^n$. The following figure shows a city with half a dozen streets creating 64 quarters.



Order in chaos: Varignon's theorem

Varignon's theorem is a celebrated geometric plane theorem which is noteworthy for the surprising phenomenon that it demonstrates. "A mathematical problem for demonstrating", as George Pólya would say. It is included here to show two fundamental principles, namely: a demonstration that does not manage to explain the phenomenon is not sufficient, and secondly, that which is pursued by the creative mathematician, unlike the logical, formal mathematician, is to comprehend the phenomenon, so a comprehensive justification becomes essential. Or that sometimes demonstrating is not explaining, which is the same thing.

Choose four arbitrary points P, Q, R, S on a plane and join them with segments to form a quadrilateral. Mark the middle points A, B, C, D of the four sides and join them to form another quadrilateral inside the first one. What do you see?



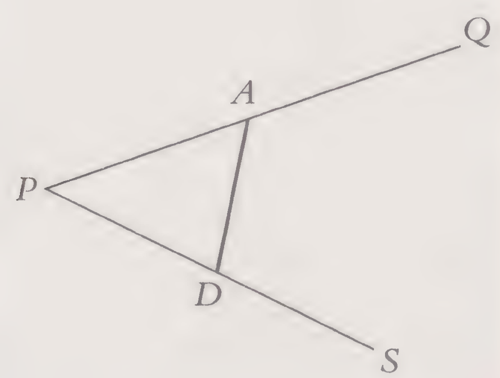
Repeat the process with four other points and your observation will be the same. We have found an unusual phenomenon. It seems that the nature of geometry does not listen to reason. However wayward the original quadrilateral, the conclusion is always the same:

The quadrilateral with vertices on the middle points of the other quadrilateral is a parallelogram.

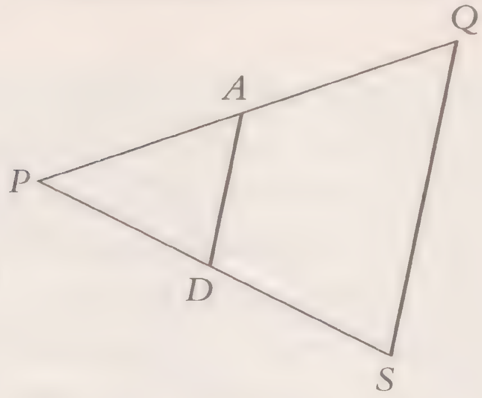
Order among chaos. In a situation like this, the basic objective is to explain the phenomenon. Maybe a demonstration could explain it... or maybe not. Let's start by looking at a vectorial and algebraic demonstration of the problem. We have to demonstrate that points A, B, C and D in the middle of the sides of quadrilateral $PQRS$ describe a parallelogram. This is the same as saying that vectors $\overrightarrow{AB}$ and $\overrightarrow{DC}$ are equal, that is, they have the same components. As $P(p_1,p_2)$, $Q(q_1,q_2)$, $R(r_1,r_2)$ and $S(s_1,s_2)$, we can find the components of the first of these vectors, which are identical to the second:

$$\overrightarrow{AB} = \left(\frac{r_1 - p_1}{2}, \frac{r_2 - p_2}{2}\right) = \overrightarrow{DC}.$$

Therefore, we have demonstrated the theorem. However, does this demonstration explain the mystery of the phenomenon? No, what we have here is an example of how logic proves, but does not explain. Let's go back to the start and stop to observe one part of the shape:

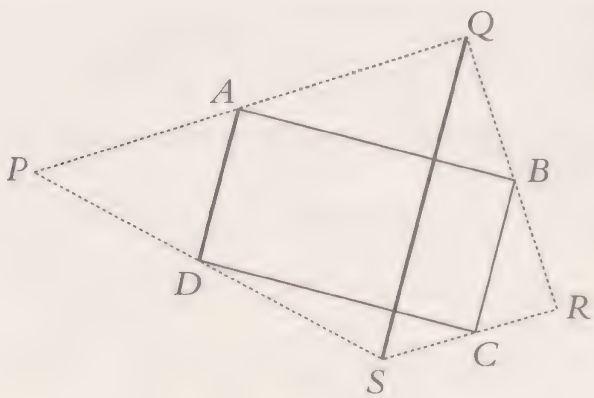


Perhaps this design is familiar, although from a context other than analytical geometry. It may be revealing to draw an extra line, the only one we can add to complete the drawing:

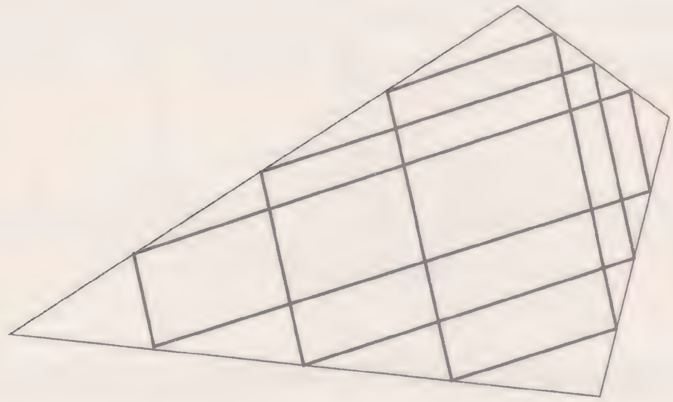


We see two triangles appear, *APD* and *QPS*. As *A* and *D* were chosen in the middle of *PQ* and *PS*, respectively, we can be sure that *AD* is parallel to *QS* and that it is exactly half its length. The latter conclusion goes by the name of the middle parallel theorem, and it is worth a special mention, as it is not as obvious as it seems.

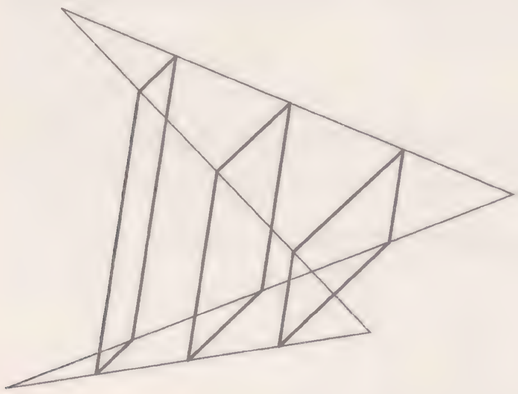
Similar reasoning in terms of vertex *R* of the original shape would lead us to conclude that *BC* is also parallel to *QS*. As *AD* and *BC* are both parallel to *QS*, the conclusion is that they are parallel to one another, and so quadrilateral *ABCD* is a parallelogram. The theorem makes complete sense in the geometric context, and the demonstration directly related to the Thales theorem shows what has to be revealed.

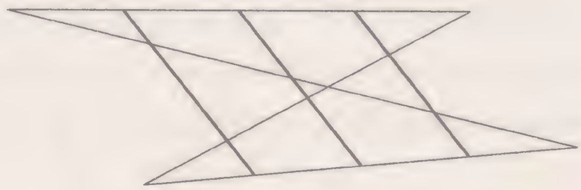


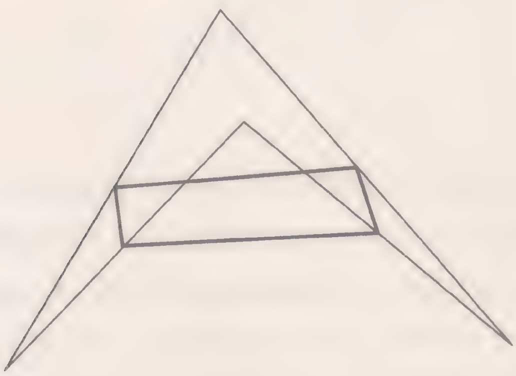
As creative mathematicians, we should not stop there. Paul Matussek, quoted in the first chapter, has said that the creative mind does not stop thinking. For example, a direct consequence is that the angles of the vertices of parallelogram *ABCD* are those that form the diagonals of quadrilateral *PQRS*. Other questions can arise: what if instead of taking the midpoints, we divide the lines into three, four, or even more parts?

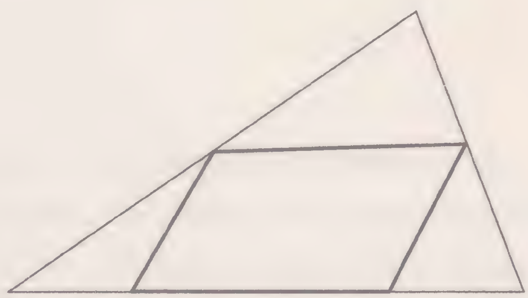


This is where software comes to our aid. Far from just being a tool, computer processing not only eases the visualisation of the phenomenon, but it can also inspire new questions and situations. The following figures have been created with a dynamic geometric program which allows the movement of any vertex of the original quadrilateral. Moving them produces extraordinary quadrilaterals and parallelograms, which go against the most conservative of conceptions:









We cannot help but see some of these figures as representations of three-dimensional polyhedrons on a plane. Varignon's theorem abandons the plane and goes into space. With a click and a drag, modern technology has broken the implicit

boundaries of the initial proposal. As a consequence, new questions arise: is the theorem still valid if the sides of the original quadrilateral intersect? What happens if one of the vertices of the original quadrilateral is on top of one of the other three and the quadrilateral has become a triangle? What is that triangle like and what relationship does it have with the parallelogram inside it? Under what conditions can the theorem be generated in space, substituting the terms quadrilateral and parallelogram for polyhedron and parallelepiped, respectively?

The powers of two are not the sum of consecutive natural numbers

The year 2000 was the World Mathematical Year. Numerous congresses were held and a multitude of mathematical activities were carried out. The roundness of the year was the inspiration of a question which the author of this book had never considered:

Is 2,000 the sum of consecutive natural numbers?

This led to a theorem on numbers unknown to the author and his then work colleagues. The manuscript for this book was first drafted in 2010, and that constitutes a number which is round enough to be able to ask the question:

Is 2,010 the sum of consecutive natural numbers?

It is not the sum of two consecutive numbers:

$$2{,}010 = 1{,}005 + 1{,}005 = 1{,}004 + 1{,}006.$$

But it is the sum of three and four consecutive numbers:

$$2{,}010 = 669 + 670 + 671.$$
$$2{,}010 = 501 + 502 + 503 + 504.$$

Is it possible that all natural numbers can be expressed as a sum of other consecutive natural numbers? Evidently, all natural numbers are the sum of one consecutive number: the number itself. Let's write the sum of k consecutive natural numbers:

$$(n+1) + (n+2) + \ldots + (n+k) = k \cdot n + (1 + 2 + \ldots + k).$$

The sum in brackets was already calculated in the previous chapter:

$$1+2+...+k=\frac{k(k+1)}{2}.$$

Returning to our case:

$$(n+1)+(n+2)+...+(n+k)=\left(n+n+...\overset{k,times}{}...+n\right)+\left(1+2+...+k\right)$$
$$=kn+\frac{(1+k)k}{2}=\frac{k\cdot\left(2n+k+1\right)}{2}.$$

On the one hand, if k is even, $2n+k$ is also even, and $2n+k+1$ will be odd. But on the other, if k is odd, $k+1$ is even, and $2n+k+1$ will also be even.

In either case, the first or the second, there is an odd factor in the numerator. Therefore, the sum of consecutive numbers has at least one odd divisor. This means that only natural numbers with an odd divisor can be expressed as the sum of consecutive numbers. Given that the powers of 2 lack odd divisors, we have arrived at the following theory:

Only the powers of 2 cannot be expressed as the sum of consecutive natural numbers.

Listing the sums of consecutive numbers we can see where the odd factor is:

Summands		Number
2	$(n)+(n+1)$	$2n+1$
3	$(n-1)+(n)+(n+1)$	$3n$
4	$(n-1)+(n)+(n+1)+(n+2)$	$2(2n+1)$
5	$(n-2)+(n-1)+(n)+(n+1)+(n+2)$	$5n$
6	$(n-2)+(n-1)+(n)+(n+1)+(n+2)+(n+3)$	$3(2n+1)$
7	$(n-3)+(n-2)+(n-1)+(n)+(n+1)+(n+2)+(n+3)$	$7n$

If the number of summands n is odd, that factor is n; if the number of summands n is even, the factor is $2n+1$ – an odd factor, in both cases.

CARL FRIEDRICH GAUSS (1777–1855)

This German mathematician, who was born in Brunswick and died in Göttingen, was a child prodigy. Gauss was unsure whether to study philology or mathematics. In early spring 1796 he chose the second option. Mathematicians are grateful, as he ended up becoming one of history's greats. His decision was undoubtedly influenced by the fact that on that same spring day he managed to construct, following Euclidean laws, the seventeen-sided regular polygon using only a ruler and compass – the first person to do so. As a mathematician, Gauss would achieve enormous success, but nothing made him as prouder than this early achievement. So much so that he wanted the polygon to be engraved on his gravestone, something which the stonemason rejected, saying it was too difficult and it would end up looking like a circle.

A portrait of Gauss. The German mathematician demonstrated that the regular seventeen-sided polygon could be constructed with a ruler and compass.

Chapter 4

Cultural Interaction and Creativity

So far we have dealt with the most common aspect of mathematical activity: an individual responds to phenomena and experiences by pondering explanations for them and offering mathematical reasons. We have not taken an in-depth look at the cultural and social aspects, despite the fact that in the first chapter they were declared fundamental for the development of mathematical knowledge.

Mathematics is developed within a specific society and culture, and it is mainly that context that determines its advance in one direction or another, both inside and outside of academia. Thus, cultural factors influence the creative aspect by lending more importance to some problems than to others, and even ignoring certain things that could be significant in another culture. Ethnomathematics is a branch of the science which studies the development of vernacular mathematical knowledge, that is, mathematical ideas from a specific cultural group. It tells us that in different parts of the world geometric figures are counted, calculated and conceived differently and problems are resolved with different procedures. On the one hand, this tests the creative character of each culture; on the other, it provides a new point of view on the same problem.

There follows a discussion on the identification and management of mathematical knowledge drawn from the author's everyday experiences outside of his academic day job and beyond his Western background. Please indulge the openly personal tone which runs through the story.

So far we have talked about heuristics in culture itself, inside or outside of the field of mathematics. We will now go beyond our cultural environment to see how mathematical creation can be inspired and managed by social and cultural aspects. We have already said that the first step in mathematical creation is formulating questions on phenomena, and what could be better than leaving our homeland to observe and experience ever more new things?

The intercultural journey

Mathematicians do not travel much. We are not talking about travelling for work reasons where the luggage for the journey is the work itself. We are talking about travelling in order to discover new things, other people, other cultures and customs, other ways of living and thinking - an activity where there is no place for organised trips, on which tourists also take a large part of their everyday lives with them in order to guarantee a few minimum luxuries, private transport, a guide and interpreter of the local language and the company of familiar people.

An intercultural journey is none of these. In a journey like this, we change context, society and culture. We live, temporarily, in a way that is as similar as possible to the people of the place we are visiting, we conduct ourselves as if we were locals, we eat the same things in the same place as them, and we stay in hotels where they stay when they travel around their homeland.

During trips like these we realise that each culture has developed what we call a particular way of life: each country has its own language, beliefs and rituals, its class system, its social and political organisations, its values, its gastronomy, architecture, art, music, literature and many other things that constitute its knowledge, and ultimately its culture. What the traveller is seeking is to share and contrast his points of view with those of others. In this way not only will he get to know the people he is visiting, but he will also find out more about himself and, therefore, his culture.

Extramural mathematics

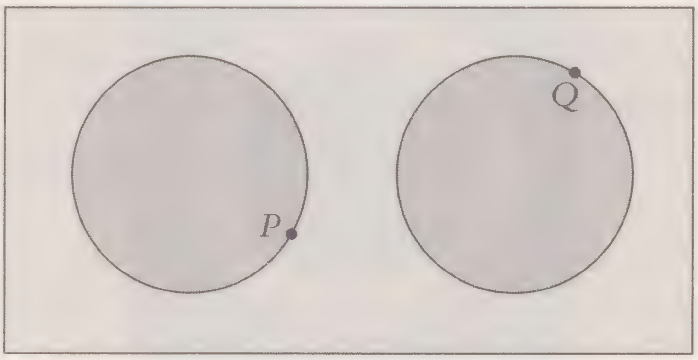
Who wants to think about mathematics when travelling? The boat carrying us glides down the Ganges while we contemplate the columns of smoke emanating from the bodies burning on the shore. Lying on the sand we see the backlit, star-shaped, swaying silhouette of the crown of a coconut palm, rustling in the breeze. Sitting on the floor of the temple we are captivated by the endless sequence of events. The priest who blesses the offerings among sandalwood smoke, the colours of the assistants' clothes, the music of the gamelan with its reiterative motifs, the sculptures that surround the enclosure, the temporary decorations made of bamboo and interlaced leaves, the baskets of exotic fruits. Can anyone think about mathematics while experiencing all of this?

The answer is no. However, when you spend enough time in a foreign country you get used to the exotic and the new. What previously seemed strange, now becomes normal. At this point you can do what you did at home, only elsewhere. Once over the first stage, and if the journey is going to be long, you can take something from its context to think about during times of boredom and while waiting.

The first time in my life that I slipped a mathematical conversation into a trip of this type was in the city of Baños, Ecuador. I entertained myself for a while there with a German friend interested in the theorems of eigenvalues.

The second was years later, when I took notes and a mathematics book to Indonesia to prepare for my classes. It was going to be a long trip, and I expected to stay several weeks in one place. Also, several works of literature and music helped to recreate a working environment very similar to that of home. I found it strange at first, but in the end I got used to doing mathematics in the tropics. It was then that the idea of the following optimisation problem occurred to me, inspired by the near-perfect circularity of the Isle of Ternate and nearby Tidore, separated by a strait of just two kilometres, in the Maluku archipelago.

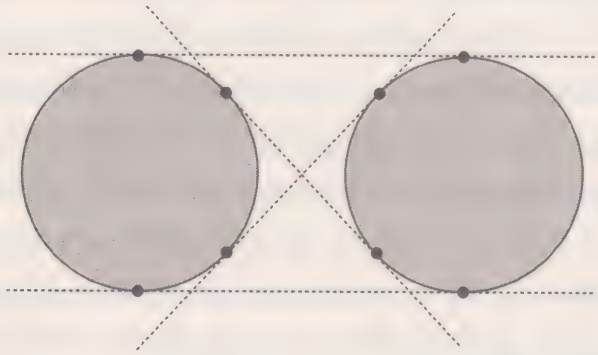
We have two circular islands. On each of them there is only one road, which completely circumnavigates them following the coast. One person is at point P of one of the island's roads and has to get to point Q on the other's road, as shown in the diagram. What is the shortest route?



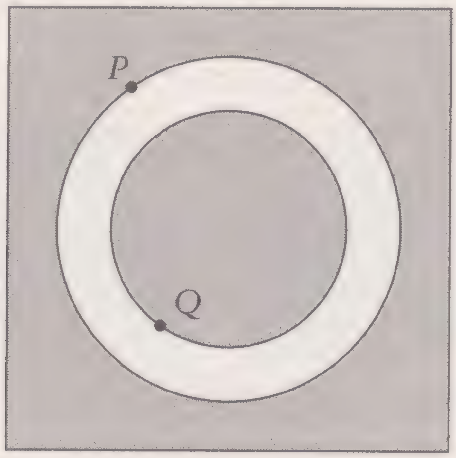
The only possible paths on each of the islands are circular, that is, circular arcs. On the other hand, they can go from one island to the next sailing in a straight line. A good way of starting is to remember Pólya and reduce the problem down to simpler cases, then increase the complexity – and perhaps even exceed it to create a general theory:

1. Both islands are points.
2. One of the islands is a point.
3. They have the same radius.
4. They have different radii.

The resolution of the problem goes through these four basic points, those determined by the common tangents of both circles, as shown in the figure.



In trying to resolve it, another basic question arises: which points are closest following an exclusively marine path? What if one island was inside the other, separated by a circular lake?

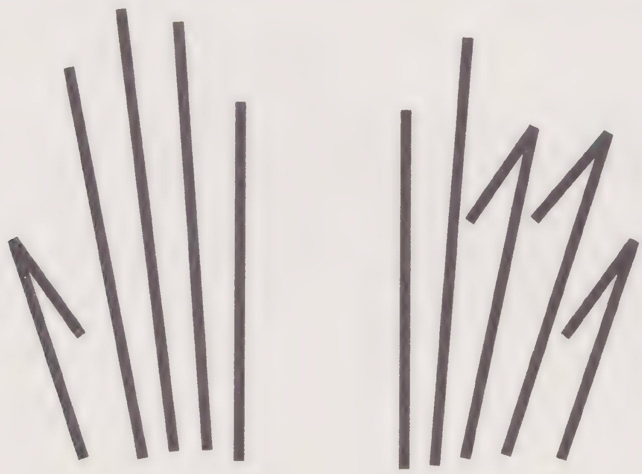


Finger multiplication on the beach

My third mathematical experience outside of my cultural comfort zone was on the beach in Padangbai, Bali. There we had met a teacher at the school in a nearby small village and his seven-year-old daughter. He wrote sums in the sand with his finger for the girl to solve. What surprised me was that she used her fingers to

multiply. I remembered a book by Georges Ifrah which was on a shelf at home, 3,000 km away, in which he talked about digital multiplication in different parts of the world, but I could not remember if Indonesia and, more specifically, the island of Bali, was mentioned.

To multiply, for example, 6 by 8, the girl counted to 6 on her closed left hand, so she ended up with one bent finger, and the other four straight. Then she did the same with the right hand, with which she was counting to 8, leaving her with 3 bent fingers and 2 straight ones, as shown in the following diagram.



She obtained the result by adding a ten for each bent finger, $10 \cdot (1+3) = 40$ to the product of the straight fingers, $4 \cdot 2 = 8$. The total is $40 + 8 = 48$.

Who would have thought of this system? Why did it work and what is achieved by using it? It is impossible to answer the first question, as its tracks are lost in time and dispersed through space. However, in response to the second question an application of the method is as follows:

$$\begin{aligned} &(10-a)(10-b) \\ &= 100 - 10a - 10b + ab \\ &= 100 - 10(a+b) + ab \\ &= 10[10-(a+b)] + ab. \end{aligned}$$

In other words, $a+b$ is the number of straight fingers, and $10-(a+b)$, the number of bent fingers. These should be seen as tens, as they are multiplied by ten. Finally, a and b are the bent fingers on each hand. The case of $6 \cdot 7$ corresponds with $a=4$ and $b=3$. For $8 \cdot 8$, we will have 3 bent fingers on each hand, that is, 6 tens (60), and 2 straight fingers on each hand, or, $2 \cdot 2 = 4$ units. Therefore the result is $60 + 4 = 64$.

What is achieved by this? It reduces the multiplication of numbers greater than five to numbers less than five. Using this finger (or digital) calculation it is not necessary to learn the times tables up to 10; knowing them up to 5 is enough.

When I got back home I looked for the book on the history of numbers. Ifrah cited various parts of the world where finger multiplication is used: "We can still find traces of this type of technique in India, Iran, Syria, Serbia, Bessarabia, Wallachia, the Auvergne and North Africa." The places cited go as far east as India, but not Indonesia. It was the first time in my life I had found something that was not documented in a book. The people of Bali are Hindus; the teacher and his daughter's finger calculation was surely an inheritance from Indian culture.

FINGER MULTIPLICATION OF NUMBERS GREATER THAN 10

The trick of finger multiplication is the remainders of a division by 5. To multiply in this way, we use our fingers. As we have five on each hand, we end up with as many open or closed fingers as the remainder of the division by 5 of the number in question. To multiply 13 by 14 we count the ones on each hand. We will see 3 closed fingers on one hand and 4 on the other, which correspond to the remainders of the divisions of 13 and 14 by 5: $13=5\cdot 2+3$; $14=5\cdot 2+4$. What do we have to do with them to get the result? Algebra provides us with the answer:

$$(10+a)(10+b)=100+(a+b)\cdot 10+ab.$$

That is, adding as many tens as closed fingers $(3+4)$ and their product to 100:

$$13\cdot 14=100+(3+4)\cdot 10+3\cdot 4=100+70+12=182.$$

Toraja carvings: can they be done without mathematics?

This event changed my perspective. From that point I started to look at things in another way. Among those things was something very special – the ornamental carvings in Toraja architecture on the island of Sulawesi, also in Indonesia. The first time I saw them I only looked at them as works of art and as an element of their culture. The traditional architecture is both characteristic and symbolic of the Toraja people. The

houses and the granaries for storing rice are made of wood and are built on thick pillars. The lower part is normally used to store various belongings and livestock. Above it are the living quarters, the walls of which are formed by an ensemble of wooden panels. The roof that covers the construction is saddle-shaped.

Identical and equidistant repetitions of geometric motifs and the division of circles into 16 equal sectors.

A multitude of designs are carved onto the four façades with scenes central to the Toraja's conception of the world, the cosmos and society. The designs are inspired by natural shapes, which are then surrounded by abstract geometric patterns.

After the mathematical experience on the Bali beach and after the realisation that this place did not appear in Georges Ifrah's book either, I went over some photographs of Toraja architecture. In them I continued to see the art and symbolism, but I also noticed straight, parallel and perpendicular lines, circles, spirals, transformations, reflections, symmetries and other geometric elements drawn with great care. Was all this possible without mathematics?

Geometry is always present in the art and decoration of the Toraja people of Sulawesi.

Observing the creation of Toraja carvings

When contemplating the carvings, the Western mathematician relates them to Euclidean geometry and considers it very probable that they are based on Euclidean concepts and procedures. This is an interpretation that is implicit in the description of these shapes' geometric elements. The terminology with which the mathematician describes the carved figures comes from his own mathematical culture. However, none of those terms have been uttered by the craftsmen responsible for the carvings.

As soon as mathematicians had the opportunity to make contact with the craftsmen and see the carving process in action, they noticed that some procedures and tools were not Euclidean at all. The engravers carried out their designs with great precision: in the vast majority of cases each carving was made using a reticle, or network of fine lines, and some designs were sketched. In both cases the designs were well defined. Their lines, drawn with utter precision, were used as a reference for the shapes that were to be carved later.

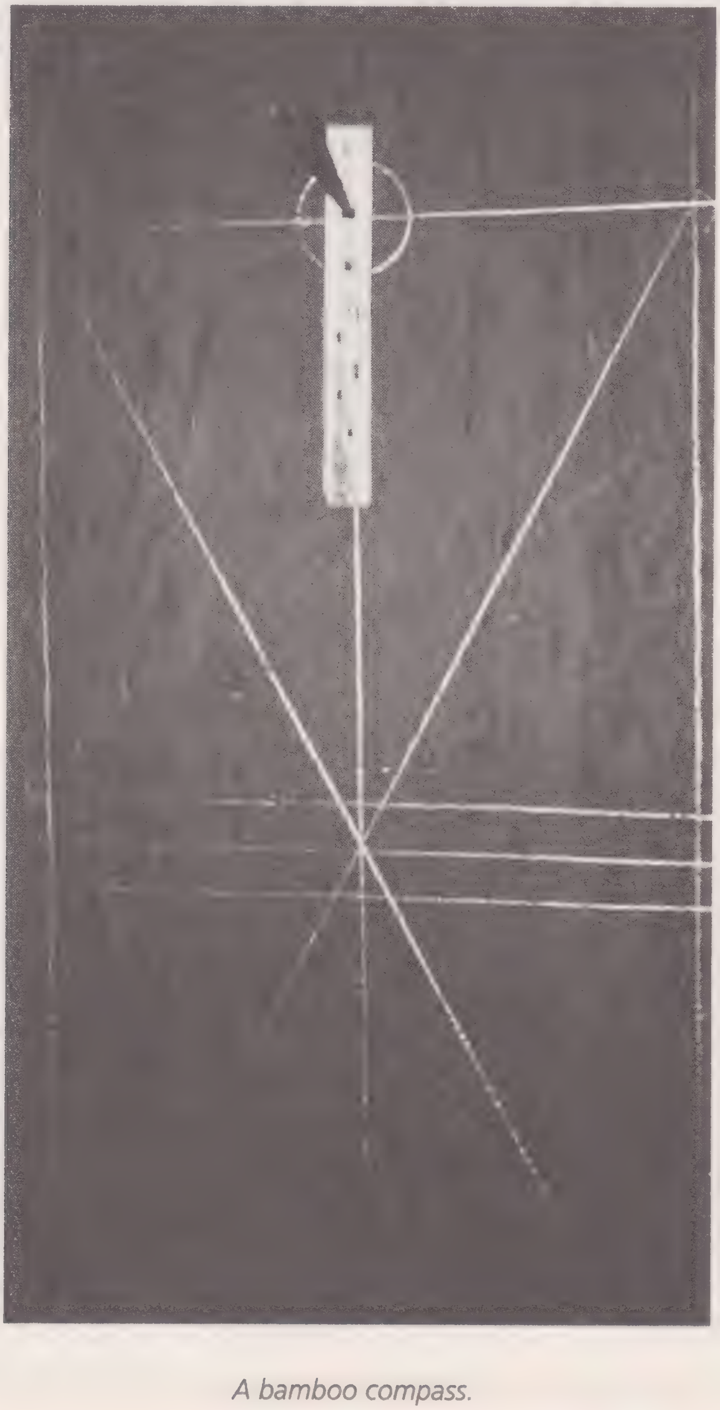
In the process of drawing the reticles it became evident that the parallel and perpendicular lines that formed them were not drawn following the propositions in Euclid's *Elements*, but according to visual estimation methods that seemed much less precise than their final appearance would indicate. It was the parallel lines and circles that were rendered with the greatest precision. These were drawn with bamboo compasses, which were effectively the same as Western ones, but equipped with two points for marking the wood.

A Toraja craftsman carving self-parallel spirals.

A handmade spiral with its eye on an intersection of the network.

No measurements were taken and no explicit calculations whatsoever were carried out. A bamboo stick was used both for taking measurements and transferring them to any point of the carving panel in order to draw rectilinear segments.

One problem was more noticeable than the rest. Before drawing the lines for the reticle in the rectangular space to be carved, the craftsman had to mark out points on the edges of the rectangles. The two, three, four, six or eight equidistant points were marked through a system of trial and error, using a bamboo stick with given



A bamboo compass.

radii marked as holes (see below) or with the more familiar form of a compass. Surprisingly, the craftsmen hit the correct point with great precision and with astonishing speed.

Eureka!

My first interpretation of this phenomenon was to think that the craftsmen wanted to divide the sides of the carving space into two, three, four, six or eight equal parts and that it was those points that they would use as the ends of the lines that would form the carving's network. Their solution was not what I would have done: I would have measured each side of the area and divided its value by the number of parts in order to later mark the corresponding points on the side. The engravers had rulers with millimetres and calculators, but they did not use them. However, their solution was as efficient and precise as they needed. It was clear to see.

Then I asked myself: "How would I do this using their tools?" I imagined myself taking an initial visual estimation of a half, a third or a quarter of the segment, which I intended to divide into two, three or four equal parts. I would mark that estimated half on the bamboo stick and then copy it onto the side of the rectangle which was to be divided. I would then move the stick until its end coincided with the mark made on the side of the area. If the estimation of the half was correct, the marked point of the stick would coincide with the edge of the area.

That is what the craftsmen seemed to be doing. But the key to the matter was right there: if they did not coincide it meant that the estimation was incorrect. So, how can the error be corrected. They reacted so quickly that it was difficult to see what their references were. It seemed that they worked by trying several times until, in the end, they got the result they were looking for. Was the problem being solved by chance? If so, how could the speed with which they obtained the solution be explained? Or perhaps I had undervalued chance and it was much more productive than I had imagined.

The next morning we took a bus. The road ahead of us was long, and it was a bright day. So, during the first part of the trip I spent my time contemplating the comings and goings of life in the rice paddies and the places that we travelled through. I felt great, and my mind was alternating between reality and fantasy. One minute I saw what I was looking at, and the next I saw what I was imagining. Suddenly, a question came into my head: how can the error be corrected in order to give an exact result?

Once again, I imagined myself in the place of the craftsman, the bamboo stick in one hand and the pencil ready to mark the suitable point in the other. I made the first estimation of the half, marked it on the stick and copied it to the edge of the carving area. Then I slid the bamboo towards the end. The mark did not coincide, the estimation was incorrect, but... Eureka! Why did it not occur to me before? In order to correct it, what had to be done was to estimate the half of the error made and subtract it from the first estimation, or add it to it, according to whether it was an over- or underestimation. This would give me the solution because that sum or difference in halves would be half of the side of the area. And if it was being divided into three parts, the same applied: I would have to correct the error by finding a third of the error made. If, after carrying out the process twice, I did not obtain the correct result, I repeated it again.

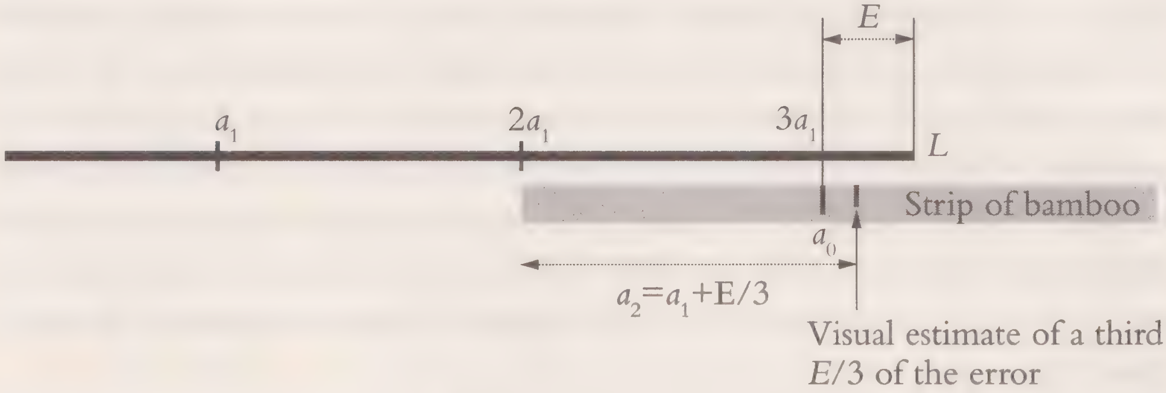
A craftsman correcting the error made in an estimation.

Let L be the length of a line segment that we want to divide into three parts. We make our first visual estimation a_1 of what we think is a third of the segment. We mark on it three consecutive points corresponding to lengths a_1, $2 \cdot a_1$ and $3 \cdot a_1$ (see the diagram below).

If the last mark coincides with the end of the segment, the partition is correct, and the problem is solved. Otherwise, the error committed, E, has to be corrected. How? Simply by finding a third of it, $E/3$, and adding it to or subtracting it from the initial estimation a_1, depending on whether it was an under- or overestimation:

$$a_1 \pm E/3.$$

The result will be the new estimation a_2 with which we will start the process again:



The estimates would undoubtedly form a sequence approaching the solution, as estimating a half or third of a very small segment like that of the error made is much easier than doing it on a long segment such as the original carving area. This process and a human's acute visual perception would lead me to the required result. As was the case.

Do they think of it like I do?

Mathematical analysis of the process confirmed my expectations, the convergence of the series towards the solution. This led me to ponder a question that was perhaps even more tricky, but fundamental to my investigation into the mathematics of the Toraja craftsmen: do they think of it like I do? I could not go and directly ask them such a question. It was essential to get them to explain what they themselves were thinking when they solved the problem.

Some spoke a little Indonesian, but most of them communicated in the local Toraja language. Until then I had used English interpreters, but I had noticed that, sometimes, instead of translating what the craftsman thought or said, they offered me a personal interpretation of his words. In a matter such as this I could not risk not knowing the truth. As I already knew a little Indonesian, I decided to learn a little more and undertake a direct conversation with the creators of the carvings. The fact that I already knew some of them could make social interaction a little easier for me. Cultural interaction had already begun.

Non-Euclidean division of a segment into equal parts

It was hard work finding a statement from one of the craftsmen confirming that they considered the division of the segment into equal parts exactly as I had imagined. Their means of working indicated as much, but given that I wanted a verbal statement of their thoughts I ended up acting as an apprentice so that they could explain how to proceed. So I prepared myself to carry out the task in situ. I took the tools they used and began to carry out the divisions on a wooden panel. I thought I was reaching my objective when, upon asking a craftsman what to do with the error in the estimation of the half, he responded: "The excess divided into two parts." I then asked what to do when dividing into three: "The same, divided into three".

I called this procedure the 'kira-kira method', as *kira-kira* is the Indonesian term for 'approximately'. The division is approximate, but not in any old way: It is the construction of a series approaching the solution which is accepted when it is less than the thickness of the pencil with which the marks were made or when the error is visually indistinguishable. It is a recurring, non-Euclidean solution, which is workable on surfaces on any plane, be they horizontal, vertical or upside-down. These are situations in which the Euclidean solution to the problem, which is the one taught in both Western and Indonesian schools, becomes impractical. The craftsmen learn from other, more experienced craftsmen, and many of them have barely been to primary school. This is a new mathematical solution, at least for Western mathematics, to one of the oldest problems. It is truly an ethnomathematical creation.

The kira-kira method is nothing more than dividing, and it is a more literal process than it may seem at first. So much so that it is possible to apply it to numerical division and see that, basically, they are equivalent processes.

How do we divide two numbers? We start by seeing how many times the whole divisor fits into the number being divided. If the value is not exact, a remainder is generated, the difference between one and the other. For example, in order to divide 13 by 5 we say that 5 fits twice into thirteen and the remainder of the division is $13-2\cdot 5=13-10=3$. What do we do next? We divide the remainder, the difference of 3, by the divisor itself, the 5. Simply to make the calculation easier, instead of taking the 3 and dividing it by 5, we multiply it by 10 and divide 30 by 5. The result is exact and it gives 6. But it then has to be divided by 10, which gives 0.6, and adding it to the above quotient: $2+0.6=2.6$. And there we have the division.

This is exactly what the Toraja craftsmen do with the bamboo stick. Only they do not calculate explicitly and they begin with a visual estimation. We could also do as they do, but with numbers. For example, to divide 2,345 by 3 we could do the following:

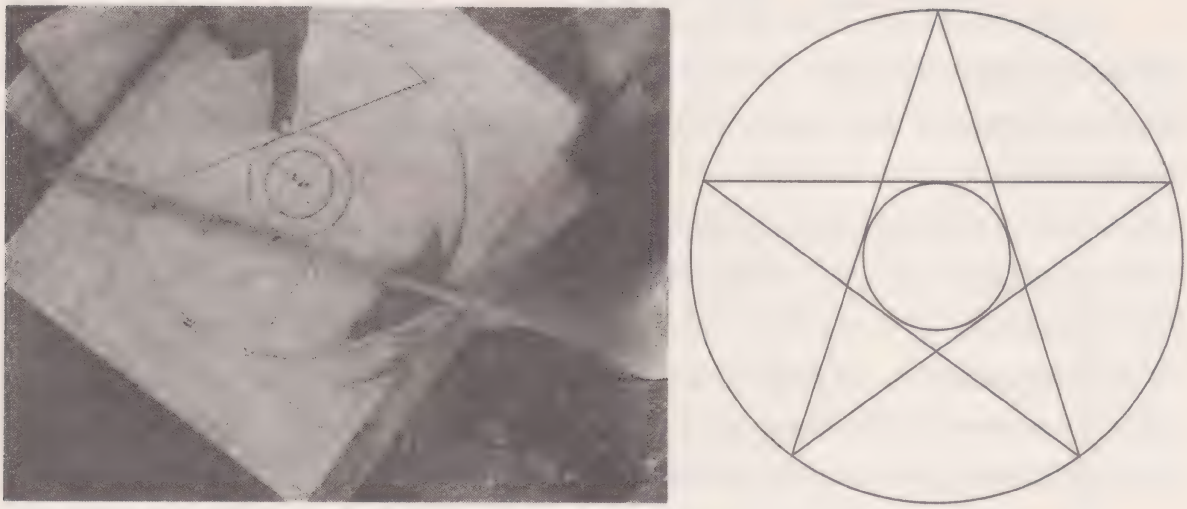
$2{,}345/3 \approx 700$	First estimation by eye: $a_1 = 700$
$700 \cdot 3 = 2.100 < 2{,}345$	First underestimation
$2{,}345 - 2{,}100 = 245$	Remainder or error committed
$245/3 \approx 100$	Estimation by eye of a third of the error which gives the second estimate: $a_2 = a_1 + 100 = 800$
$800 \cdot 3 = 2{,}400 > 2{,}345$	Second overestimation
$2{,}345 - 2{,}400 = -55$	Remainder or error
$55/3 \approx 20$	Estimation by eye of a third of the error which gives the third estimate: $a_3 = a_2 - 20 = 780$
$780 \cdot 3 = 2{,}340 < 2{,}345$	Third underestimation

The error of the third estimation ($2{,}345 / 3 = 780$) is now very small, in the order of 0.2%, compared to the exact result of 781.666... Despite this and despite the fact that it is an effective method for carrying out geometric divisions in a practical context, it is not sufficient in the context of numbers.

A new problem

A while afterwards I re-established contact with some of the craftsmen. I was surprised by the presence of an unusual geometric shape in their workshops, the regular pentagram. When questioning a craftsman, who was by now becoming a friend, about the drawing method he replied that the six-pointed star is easy, but that the five-point one is much more difficult.

It certainly is. The craftsmen inscribe it in a ring of two circles with visually determined radii. If things do not come out as expected, the error is corrected, not

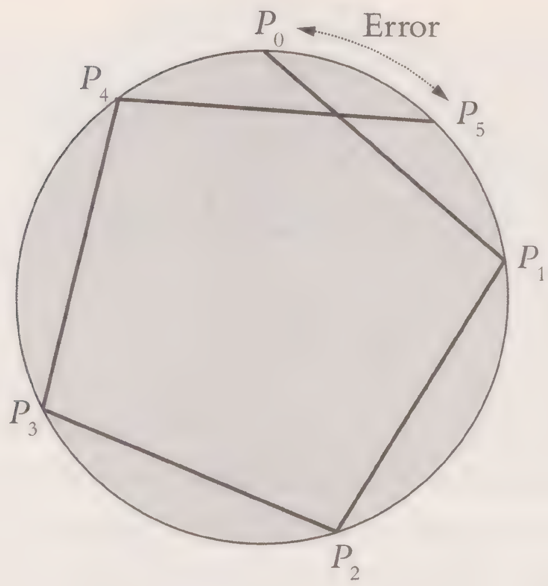


A craftsman indicates the inaccuracy of the tangents while drawing a regular pentagram.

with the kira-kira method, but visually and without a predetermined, rigorous and quantified plan to lead to the solution.

I was thinking about how I could help them to do the job more easily. It was clear that the problem was determining the five equidistant points on a circle which could be alternately joined up to create the pentagram. The problem was, therefore, equivalent to drawing the regular pentagon. The Euclidean solution to this problem was completely rejected for two reasons. Firstly, because it seemed crazy to draw a pentagon on a vertical panel of a traditional Toraja house with such a long procedure, and one which I myself find difficult to remember. Secondly, because it was not very ethical to export such a procedure to another culture. So... Eureka! Why not try to resolve the problem in the terms offered by the very culture in which it manifests itself? In other words, could the kira-kira method be applied to the construction of regular polygons? The answer is yes, but not exactly as this Western mathematician first thought.

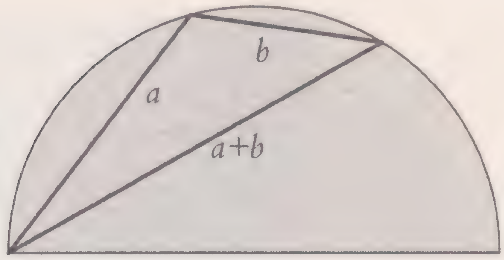
We start with a circle in which we want to inscribe a regular pentagon. We apply the kira-kira method making an initial estimate on a stick of a fifth of the circumference. Using this initial estimate we find the five points, P_0P_1, P_1P_2, ..., P_4P_5 on the circle. If the last point ends up on P_0, that is, if the last point marked meets the first one and closes the cycle, we already have the five vertices of the regular pentagon. The chords of the five circular arcs corresponding to those five marks are the sides of the pentagon. All that is left is to join them alternately to form the pentagram.



If the cycle is not closed, in other words if P_5 does not coincide with P_0, we have made an error in the estimation. At first I thought that I should correct the error by finding its third on the bamboo stick in order to add it to or subtract it from the initial estimate. However, doing it like that did not improve the result. How could it be resolved? Eureka again! It was a circle and I was still thinking about the lines. What I had to correct was an error in the arc. I should not have been focusing on the stick, but on the arc corresponding to the error. The stick marked the chords, and what I had to do was add or remove a third of the circular arc, although to do so I only had the chords of a circle provided by the bamboo stick.

The rectilinear circle

The issue was thinking of the circumference as if it were a line segment. Fixing one end of the stick at the second point on the circumference, I moved the other end towards the visual estimate of the third of the arc corresponding to the error made. The resulting distance or chord on the stick would be the new estimation. The key was in the fact that every estimation marked on the bamboo stick was a chord of a circle and... Eureka once more! The arc of the sum has to be the sum of the arcs. But this does not work by adding the chords as if they were segments, as then the resulting arc is not equal to the sum of the other two. In other words, the sum of the chords is not the chord of the sum, unless the sum of the chords is defined as that which closes the triangle determined by the other two:



I had created a recurrent and non-Euclidean construction of the regular polygons, as the procedure is applicable to the division of the circumference into *n* parts. Also, this was a new additive group that I would call the 'group of chords of a circle'. Adding the chords made sense. Closing the triangle defined by the first two meant that the arc of the sum corresponded to the sum of the arcs. The kira-kira method was powerful enough to use in the resolution of problems for which it was not created.

AN EXPLANATION FOR THE REGULAR NONAGONS IN GRANADA'S ALHAMBRA

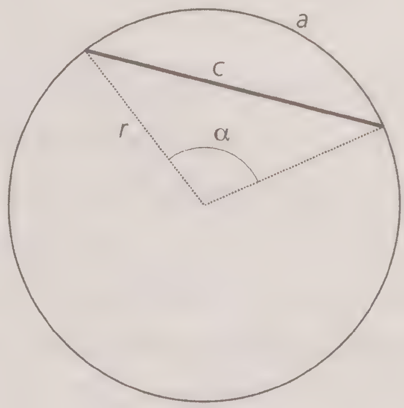
One of the Western applications of the kira-kira method is that it explains the 'trick' referred to by Pérez, Gutiérrez and Ruiz (2007) when they talk about the construction of a regular nine-sided polygon in Granada's Alhambra. Such a figure deserves the title of trick because, thanks to Gauss, we know that it is not possible to draw such a regular polygon with a ruler and compass alone. We do not know how the Arab craftsmen did it, but the circular kira-kira method provides a plausible solution to the problem. It would mean first dividing the circle into three equal arcs and then dividing each of them into three parts, always using the Toraja method. The last arc obtained will be a third of a third of the circumference, or, a ninth of it. Its chord corresponds to the side of the regular nonagon inscribed in the circle.

Notifying the Toraja craftsmen

I was uncertain whether to mention the use of the kira-kira method on the circle to the craftsmen or not. Before the pentagram problem, the craftsmen's system was a closed system, as it was capable of resolving all the problems they were faced with. However, their system was unable to solve this new problem. I was worried that telling them that the system could be transferred would mean pointing out their failings. However, what finally convinced me to do it was thinking that the new problem which tested their system had been proposed by them.

A CONTEXT FOR AN EXTRAORDINARY TRIGONOMETRIC FUNCTION

What is the error committed when using the chord of a circle as an approximation of its arc? Where a and c are the lengths of an arc and its chord, respectively, determined by an angle α in a circle of radius r:



Considering that $a = \alpha r$ and $\frac{c}{2} = r \sin\left(\frac{\alpha}{2}\right)$, we can deduce that: $\frac{c}{a} = \frac{\sin\left(\frac{\alpha}{2}\right)}{\frac{\alpha}{2}}$.

Then, function $f(x) = \sin(x)/x$ measures the ratio between the chord and its corresponding arc on the circle. We have thus created a new context for a trigonometric function, the popularity of which resided, until now, in its use as an example of an extraordinary calculation of limits. Despite not being continuous when $x=0$, the limit at that point exists and is 1. The very existence of such a limit is demonstrated by comparing arcs and chords.

A year and a half later, when I returned to the region, the craftsmen continued to produce the pentagrams as before. While I explained the circular version of their method on a circle drawn on a wooden panel, they started to understand my idea and they anticipated the results and accepted them.

Chapter 5

Mathematics for Creative People

So far we have talked about how to create in mathematics. Now we are going to shift our perspective to see how mathematics has been used in two fields that nowadays constitute the acme of professional creativity on the margins of the world of pure art, namely design and advertising.

There is no doubt about the crucial role that geometry plays and has played in design. Whenever we try to create something material and tangible geometry is involved. From the early 20th century purely geometric shapes have taken a leading role in the design of all sorts of consumer products, above all furniture and packaging. Undoubtedly influenced by an aesthetic taste for the abstract and economy of shapes, designers have worked with geometric objects to give their creations a more elegant and less jumbled appearance than in previous periods.

The world of advertising has not been excluded from that focus. In recent decades interest in science and scientific rigour has inspired advertising companies to come up with their own mathematical ideas and elements to give greater credibility to the qualities of the product they are advertising. Thus, graphs, formulae, geometric objects, symbols, numbers and calculations have taken centre stage in adverts in all media, both in print and audiovisuals.

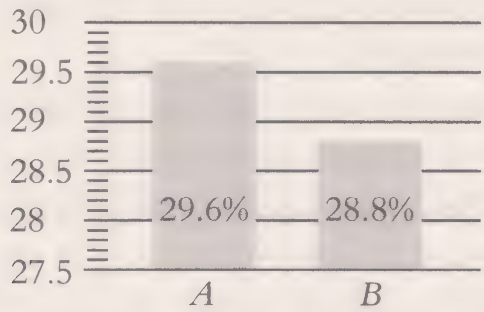
The importance of the use of mathematics in those fields is two-fold. On the one hand, the way in which both designers and publicists use mathematical ideas boosts their effectiveness. On the other, using them in contexts unlinked to the technological and scientific world opens new roads of understanding for traditional ideas or concepts that may be more accessible to the general public.

Both the worlds of design and advertising are extremely diverse. We will mathematically analyse meaningful examples from both arenas, but they are always linked to the world of advertising. Thus, when talking about a design, it will be because one of its mathematical characteristics is highlighted in an advert.

Mathematics as an advertising strategy

Biased use of proportionality

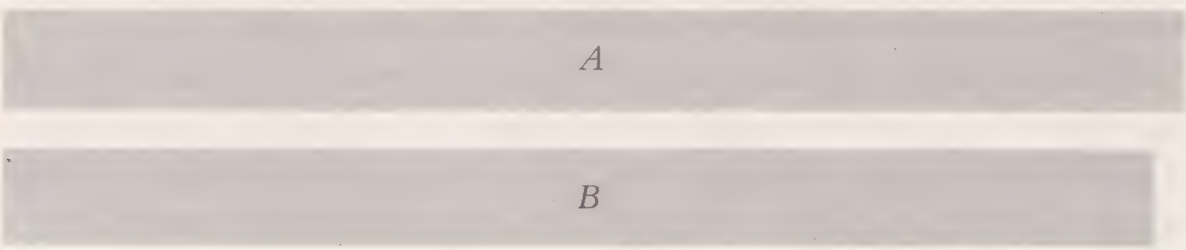
The constant battle for ratings leads radio and television stations to launch advertising campaigns which praise their own success above that of others. A common example is found in the use of graphs that attempt to give greater credibility to a statement contained in the laudatory text. A typical diagram often used to highlight the supremacy of television channel *A*'s audience over channel *B*'s is the following:



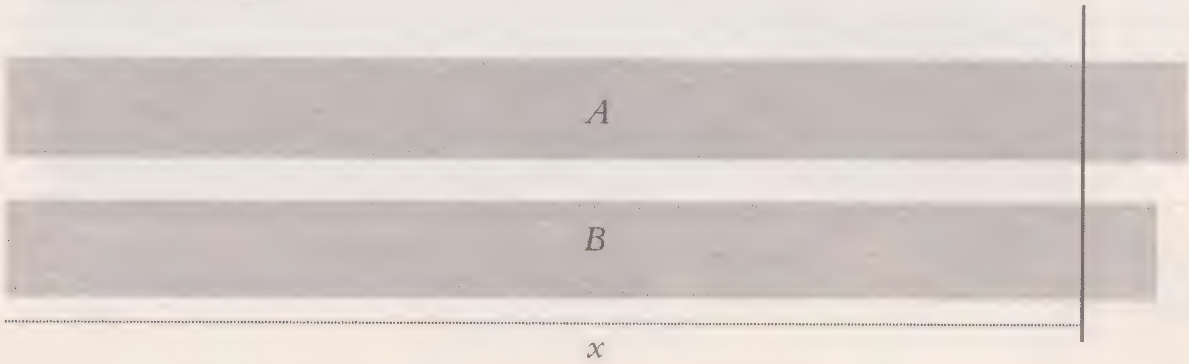
Assuming the data is correct, it is true that channel *A* has beaten channel *B*. However, the visible difference between the bars greatly exaggerates the supremacy. The bar corresponding to channel *A* is much greater in height than that of *B*:

$$A: 29.6-27.5=2.1.$$
$$B: 28.8-27.5=1.3 \Rightarrow A/B=2.1/1.3=1.615.$$

The reality is that the difference in percentage of audience is just 8 tenths, which means barely 2.8% more viewers for *A* than *B*. A true comparison with reality would be to produce the bars with their entire lengths:



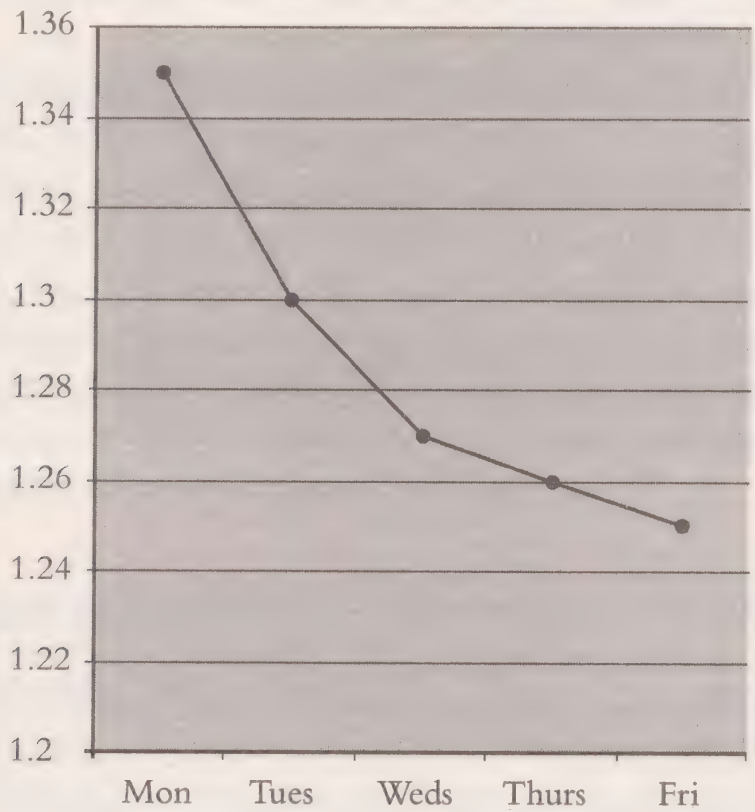
By cutting the rectangles where we wish, we can infinitely grow the apparent proportion between the resulting dimensions. The closer to the end of B the cut is made, the more this is true:

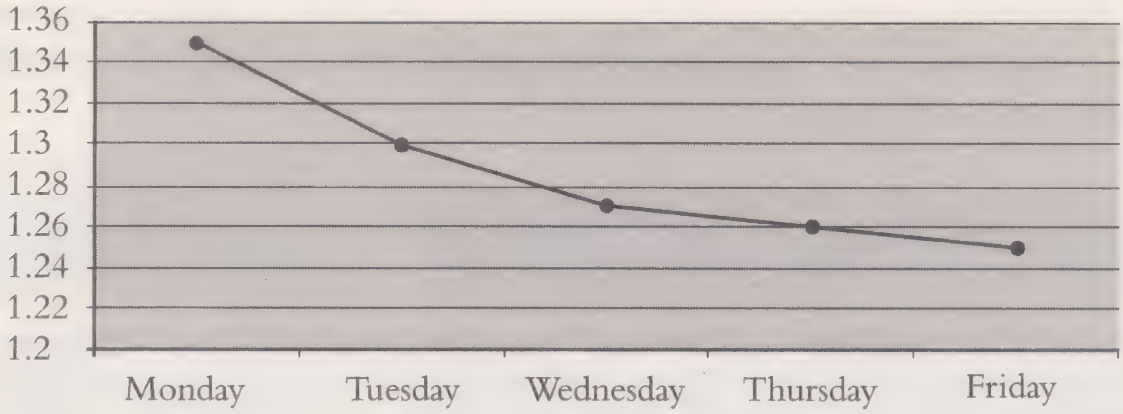


This can be taken to extremely exaggerated extremes, making the proportion as large as we like:

$$\frac{A-x}{B-x} \xrightarrow[x \to B]{} \frac{A-B}{0} = \infty$$

A similar problem appears in graphs corresponding to the variation in currency exchanges. The evolution of the exchange rate over a week can seem trivial or enormous, depending on the vertical scale used in the graph:





The pretence of illustrating proportions or differences is one of the most recurrent ideas in advertising, although, unfortunately, it is used to distort precisely what it seeks to demonstrate. Numbers and graphs provide an air of accuracy that is inherent to the field in which they are used, mathematics, but only when they are used objectively and without bias.

Probability

A few years ago there was an advert from a telephone company in which the following statement was made: "The possibility of the boss's son and you being the same person is 0.00000000001%".

A more suitable term than 'possibility' would have been 'probability'. Probability in mathematics is a way of quantifying the potential that an event will occur. The classic definition of probability establishes that this quantification is the ratio of the number of favourable cases and all the possible cases related to the occurrence of the phenomenon. For example, when throwing a six-sided die, the probability that a number, x, smaller than three will be rolled is 2/6, as there are only 2 favourable results (the 1 and the 2) out of 6 possible results (the six values of the die):

$$P(x < 3) = \frac{2}{6} = \frac{1}{3} = 0.333...$$

One way of determining the probability that a person's father is their boss is to calculate the quotient between the favourable and possible cases. As everyone only has one father, the number of favourable cases is just one. In terms of the possible cases, it would be helpful to know how many bosses there are, which is practically impossible. Let's assume that there are about six billion inhabitants on the planet

(although the number is increasing all the time). We would have to exclude the female sex (it is a father, not a mother), those who do not have children, those who are not anyone's boss and those who do not work. This would be much less than half of the six billion possible cases. An estimation would give a probability of:

$$P \approx \frac{1}{3 \cdot 10^{9}} = 3.33 \cdot 10^{-10}.$$

The probability given in the advert was:

$$Q = \frac{0.0000000000\,1}{100} = 10^{-13}.$$

Which means assuming the existence of 10^{13} bosses in a population of $2 \cdot 10^{13}$ people. That is, six hundred planets like ours.

$$\frac{P}{Q} = \frac{3.33 \cdot 10^{-10}}{2 \cdot 10^{-13}} = 600.$$

We do not know why the advertiser used the number 0.00000000001%, but, of course, they managed to make sure that it was impossible for any reader on the planet to be the son of their boss. The more zeroes there are after a decimal point, the smaller the number. Adding the % symbol reduces it by another hundredth.

This is a case of creative advertising based on magnifying an impossible event with the size of the term which represents it. Although 0% is the probability of an impossible event, the visual impact of 0.00000000001% has more impact. The use of more ambiguous expressions would not contribute anything new and would only increase confusion.

Extraordinary algebra

Some car adverts are very creative. In recent years there has been a proliferation of campaigns that stress technological, geometric and mathematical matters. The phenomenon, above all, has been seen in campaigns for high-range models.

A campaign launched a few years ago showed a vehicle reflected on a floor so perfect that it looked like glass. The resulting effect was that the car appeared to be moving across a mirror. In the image a formula and slogan could be seen.

$$1+1=\infty$$

The pursuit of perfection

Does the equation $1+1=\infty$ really have anything to do with the thirst for perfection? At first, it could appear to be incorrect to equate the sum of a number as finite as 1 with itself to infinity. An equation which brings unexpected consequences:

$$1+1=\infty \Rightarrow 2=\infty.$$

Could this actually make some sense? With the meaning that we normally give to the term infinity and with the meaning we normally give to the sum of two numbers, perhaps not. But there are other meanings for this equation that could be completely correct.

Instead of taking the full scope of the set of natural numbers, we take set X, formed by just three symbols:

$$X=\{0, 1, \infty\}.$$

If we associate the symbol 0 with zero; the symbol 1, with any finite number, and the symbol ∞, with an amount that is neither finite nor zero. Assuming this, it is logical to think that:

Nothing + anything = anything.

Finite + finite = finite.

Infinite + anything = infinite.

So the sum $1+1$ should not be infinite, but finite, and the sum of set X should be governed by the following table:

+	0	1	∞
0	0	1	∞
1	1	1	∞
∞	∞	∞	∞

This addition works as we would expect. It is a commutative addition (the table has diagonal symmetry); it has a neutral element (the zero) which, when added to another, always gives the other number as the result, and it is also associative (the order in which three numbers are added does not affect the result). Will this work in the same way changing $1 + 1 = 1$ for $1 + 1 = \infty$, as stated in the advert? That is:

+	0	1	∞
0	0	1	∞
1	1	∞	∞
∞	∞	∞	∞

This being so, the table conserves its diagonal symmetry. The zero also continues to be the neutral element. The associative property also works.

But there is something that does not work in either of these two tables: there are no opposites for the 1 or for the ∞. In neither of the two cases is there a number which, when added to 1, gives 0, and nor is there one which, when added to ∞, also gives 0. Fixing this means that in every row or column of the table there has to be a 0 at least once. Evidently, by filling the tables with zeroes the problem would be solved, but this is not the solution that 'makes sense'.

The results of the first row and the first column are undeniable, as adding zero to anything gives anything. If we define $1+1=\infty$, the table ends up as follows:

+	0	1	∞
0	0	1	∞
1	1	∞	
∞	∞		

For opposites to exist, the 0 has to appear in every row and column. If we want commutativity we have to conserve the diagonal symmetry. There are few options:

+	0	1	∞
0	0	1	∞
1	1	∞	*a*
∞	∞	a	0

+	0	1	∞
0	0	1	∞
1	1	∞	0
∞	∞	0	*b*

+	0	1	∞
0	0	1	∞
1	1	∞	0
∞	∞	0	0

There is no value of a for which the sum defined in the first table is associative:

$$1+(\infty+\infty)=1+0=1.$$
$$(1+\infty)+\infty=a+\infty=1 \Rightarrow a=1 \Rightarrow 1+(1+\infty)=1+1=\infty.$$
$$(1+\infty)+\infty=1+\infty=1 \Rightarrow \text{contradiction.}$$

The third one does not work either:

$$1+(1+\infty)=1+0=1.$$
$$(1+1)+\infty=\infty+\infty=0.$$

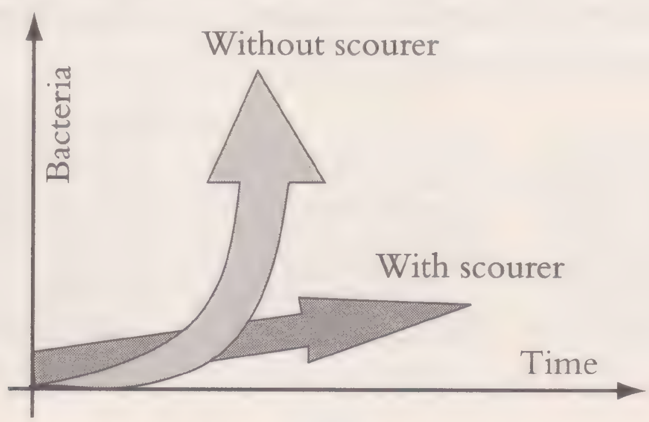
Only by taking $b=1$ can we resolve the problem in the second case. This is somewhat surprising as $1+\infty=0$ and $\infty+\infty=1$ clash with our previous conceptions:

+	0	1	∞
0	0	1	∞
1	1	∞	0
∞	∞	0	1

We have constructed an algebraic structure constituted by set $X=\{0,1,\infty\}$ and a + operation the results of which are not outside of X and have the expected coherence. Whether or not this coherence is what we assigned to the symbols with which we labelled the elements of X is another matter altogether.

Linear and exponential functions

For years a make of sponge had a slogan accompanied by a graph on its wrapping; the slogan said "...Avoid bacteria multiplying on the scouring pad...". The attached graph showed two expanding arrows in a system of coordinates in which time was related to the number of bacteria. One corresponded to the effect of using the special sponge scourer, and the other, to that of not using it.



Unlike in other cases, the correct application of the concept and the corresponding drawing of graphs deserves a mention. The population of bacteria on the sponge grows with the passing of time. The advertiser does not say they will stop increasing, just that when the product is used they will not increase by reproducing – though they may be added from elsewhere.

Let's suppose that for each unit of time $b > 1$ bacteria are created. If they multiply, after t units of time the number of bacteria will be:

$$B(t) = b \cdot b \cdot b \ldots^{(\text{t times})} \ldots \cdot b = b^t.$$

This is a function with an exponential base b. Its graph is an increasing curve with a slope that steepens continuously until it is vertical and infinite:

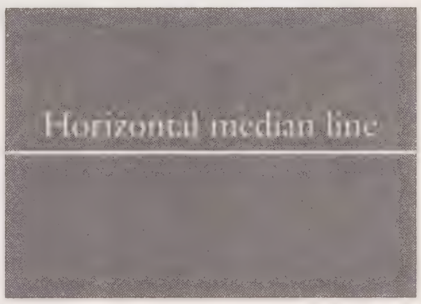
If the bacteria, instead of multiplying, are added, the formula for their population after t units of time is:

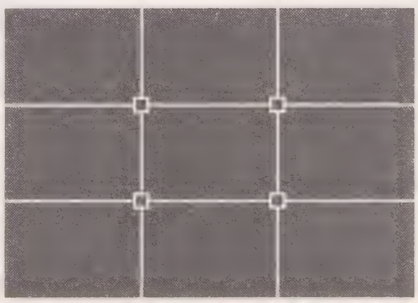
$$B(t) = b + b + \ldots^{(t\ times)} \ldots + b = b \cdot t.$$

This is an increasing linear function, with a constant slope that is exactly b, and the graph of which is a straight line. Except for the fact that exponential functions do not normally start at the origin of the coordinates, both are represented by the advertiser. Full marks! The exponential curve does correspond to the reproduction of the bacteria.

The rule of thirds

In the world of images, which is so important in advertising, there is a rule by which thirds take priority over halves. When creating an image or taking a photo it is worth avoiding having objects of interest in the middle of the image. For example, it is better to place the horizon above or below the horizontal bisector of the photograph's rectangle (the horizontal median line). Equally, if there are two objects of interest, it is better to place them on the intersections of the vertical and horizontal thirds (right). In this case, geometric proportionality is used to create an aesthetic pattern that has been validated over the years.





Mathematics for perfection

A while ago, a brand of wine launched a campaign with the message that the excellence of its product was the combined result of mathematics, nature and craftsmanship. The image in the advert showed an extremely long series of mathematical formulae, most of which did not make any real sense, in which many of the figures and letters had been substituted for images of the world of nature and craftwork related to wine production.

The end was marked by an equals sign followed by the bottled drink. By the bottle there was a slogan: "What made it perfect?"

Whether or not the formulae used were appropriate, the creativity of the advert resided in focusing on the mathematical side of the scope, thoroughness and strictness of the production process. These three aspects characterise most of the mathematical activity; together with them, mathematics contributes to the perfection of the product created.

Design with mathematics

Binary time

The binary number system only uses two figures, 0 and 1, to represent any number. As with the decimal system, the position of each figure in a number written in the binary system corresponds to a power of the base 2:

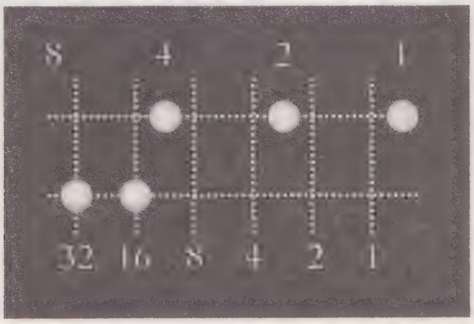
$$372_{10} = 3 \cdot 10^2 + 7 \cdot 10^1 + 2 \cdot 10^0.$$
$$101_2 = 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0.$$

Here we have the first thirteen natural numbers written in each system.

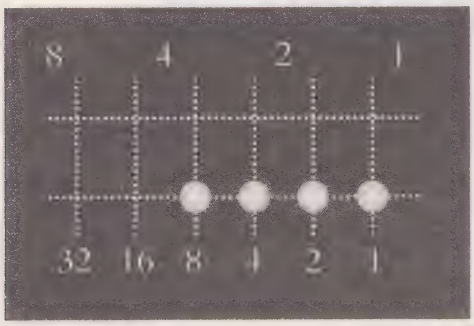
Decimal system	1	2	3	4	5	6	7	8	9	10	11	12	13
Binary System	1	10	11	100	101	110	111	1000	1001	1010	1011	1100	1101

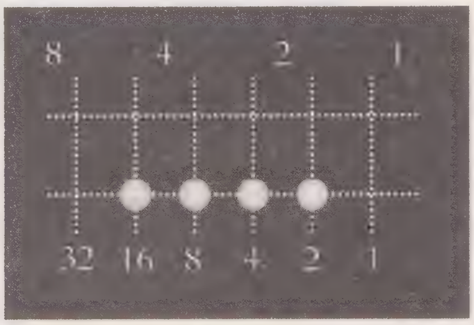
The world of design sometimes surprises us with remarkable gems. We are used to measuring time in hours of 60 minutes and minutes of 60 seconds. A watch that measures the time with a binary numbering system may be considered extravagant, but it already exists. Its face is a rectangle defined by four lines. The upper one is

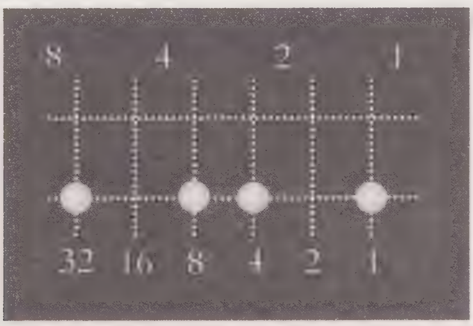
used for hours; the lower, for minutes. Between the two there are two more lines on which the values corresponding to each hour are indicated (see the diagram below). As the number of hours is a value between 0 and 12, only four digits are required to show all of them (see table above). On the other hand, for the minutes, six digits are needed to represent all the values between 0 and 60.



The difficulty lies in the fact that reading the time is indirect. The values marked on each line have to be added together in order to produce it. The diagram above shows 7 hours and 48 minutes. A quarter of an hour, half an hour and three quarters of an hour are represented as follows:



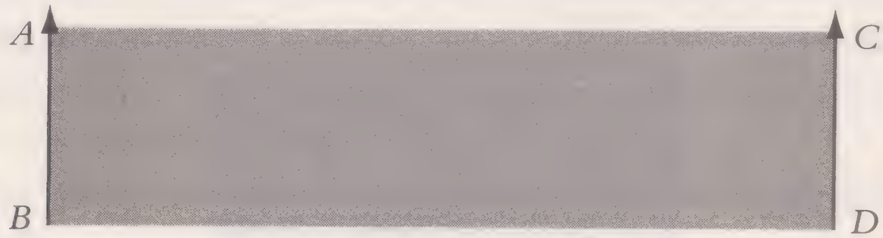




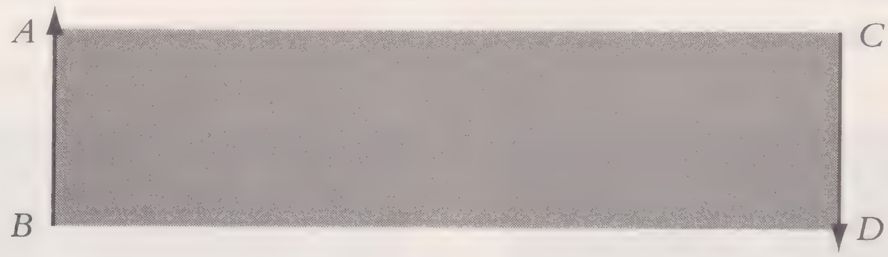
At first the watch does not seem practical, but perhaps we would get used to it and be able to read the time quickly. This is an extraordinary example of how mathematics determines the essence of design.

The Möbius strip

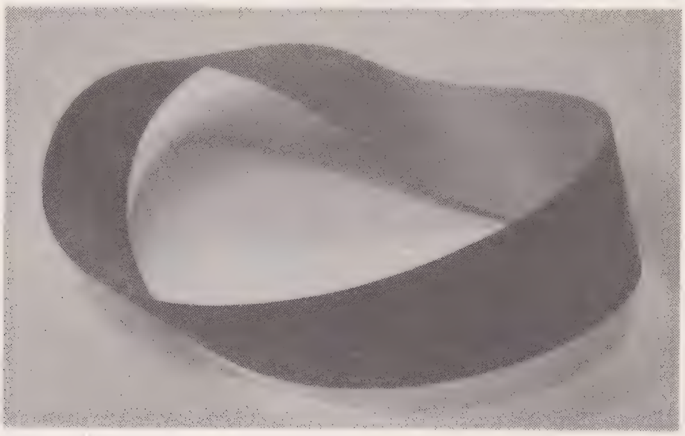
Joining together the opposite sides of rectangular strip *ABCD* creates a ring connecting sides *AB* and *CD* and joining pairs of vertices *AC* and *BD*. The arrow in the following diagram shows the way in which this connection is made.



The Möbius strip is constructed in the same way, except for one step, which provides a very different result to what we expect from its construction. The strip is also formed by joining together sides *AB* and *CD*, but in the opposite way, in other words, by connecting vertex *A* with *D* and *B* with *C*:



The result is a ring with just one face and one edge.



This is a special geometric figure used in the design of jewellery. Adverts for these rings are accompanied by texts which stress:

– The topological peculiarity of the object: "This wonderful silver ring has a unique shape with just one face and only one edge. Its shape symbolises the balance between our inner and exterior selves".
– The consequences that may arise from it are: "How is it that a small band of gold can make someone feel as if they have the world wrapped around their finger? It is perfect...".

Rings are normally cylindrical and have two faces, one exterior and one interior. The finger which they envelop is only in contact with the interior face. If the interior face envelops the finger, the exterior one envelops all that is not the finger, that is, the whole world. A Möbius ring has just one face. Therefore, that which envelops the finger is the same as that which envelops everything that is not the finger, in other words the whole world. The same can be said of the same face in terms of the exterior world. Then the Möbius ring envelops everything and envelops the world with its wearer. It does not distinguish between inside or out; everything is one. This is where the statement about having the world wrapped around a finger comes from.

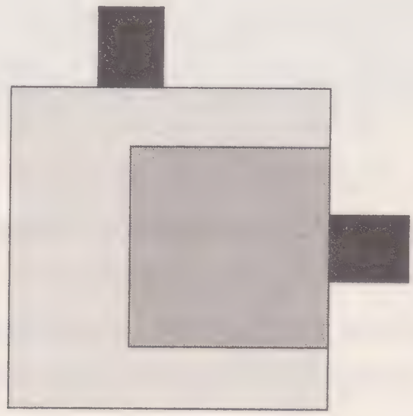
HEXOMINOES AND DESIGN

Here we have the plan of a cardboard box used to wrap a shower cap in a hotel room. It is called a hexomino because it has six equal or modular pieces.

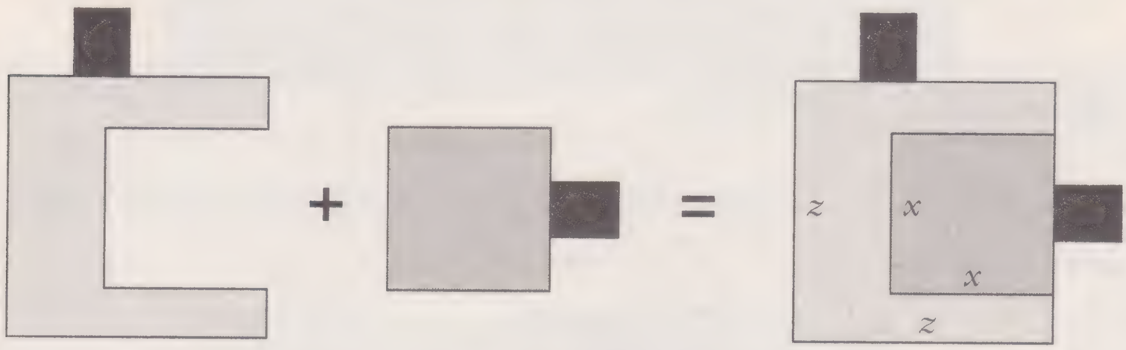
The three-dimensional polyhedron which is obtained by folding this hexomino is the cube or hexahedron. There are eleven different hexominoes from which a cube can be built.

Geometric spirit

Geometric puzzles have been used in toiletry bottle design, for both men and women. Occasionally, the algebraic expression for the composition of the puzzle in question has been provided too. This was the case with the men's aftershave and deodorant bottles of a well-known Japanese brand. The shape chosen was the square, which was broken down into two containers. One was also a square; the other, symmetrical:



The capacity of the larger one was 75 ml, while that of the smaller one was 50 ml. The advert stressed the total capacity and the way the bottles slotted together:



Can the actual dimensions of the bottles be found? The smaller bottle was 50 ml, and the larger one, 75 ml. In total, 125 ml. The fit showed that both were of equal thickness; therefore, their volumes were proportional to their areas. Taking into account that 1 ml of water is 1 cm^3, it is not unreasonable to consider that side x of the small bottle and z of the large one were:

$$50 = x^2 \Rightarrow x = \sqrt{50} = 7.1 \text{ cm.}$$
$$125 = z^2 \Rightarrow z = \sqrt{125} \approx 11.2 \text{ cm.}$$

Why do 2,000-piece puzzles not have 2,000 pieces?

We are not always free to create things with the shape or number of elements we would like. Few people will have bothered to count the pieces of a jigsaw. In fact, some people may think that it would be a ridiculous task, given that the number of pieces always appears on the box: 500, 1,000, 2,000, 3,000, 5,000, 8,000. However, manufacturers sometimes lie. Or, at least, they don't tell the whole truth.

It is true that 500-piece puzzles do have 500 pieces, but 2,000 piece ones do not have 2,000. And there is no need to count them to check this. Some manufacturers include this detail in their technical specifications, but not the reason for it. All puzzles are rectangular. Their pieces, although different frmo one another, are cut from a rectangular base on which protuberances and openings have been created. In order to create a 2,000 piece puzzle we need two integer numbers, one for each side of the rectangle, with a product of 2,000. Given that $2{,}000 = 2^4 \cdot 5^3$, the options are:

$$1 \cdot 2{,}000 = 2 \cdot 1{,}000 = 4 \cdot 500 = 8 \cdot 250 = 10 \cdot 200 = 16 \cdot 125 = 20 \cdot 100 = 25 \cdot 80 = 40 \cdot 50.$$

The proportions between the length and width of these distributions must be such that it produces a balanced rectangular distribution. Once constructed, the puzzle must not look like a strip, instead it must be similar to piece of A-series paper. This means that the proportion between the length and width is, approximately 1.4. But the rectangles derived from the divisors of 2,000 are either too square or too long:

$$\frac{50}{40} = 1.25;$$
$$\frac{80}{25} = 3.2.$$

Instead of 2,000 pieces, puzzles tend to have 1,998. We can see the reason by breaking down 1,998 into prime factors and seeing that the product of two of its divisors give a rectangle that better fits the desired format:

$$1{,}998 = 2 \cdot 3^3 \cdot 37 \Rightarrow \frac{2 \cdot 3^3}{37} = \frac{54}{37} \approx 1.46.$$

Here we have an extraordinary case in the breakdown of a natural number into prime factors which contributes or impedes, but always justifies, certain designs.

Make-up removal with Pythagoras

Wiping make-up off your face is a task that is normally done using small pads or tissues designed for the job. Each manufacturer makes its product in a particular shape, but some use very mathematical designs that venture beyond a square, rectangle or circle.

The following diagram reproduces the design of a few make-up removal sponges. It is an overhead view, but it should be understood that the sponges are three-dimensional, with a height of about two centimetres. It shows four pieces which fit together like a puzzle.

More specifically, this is the puzzle which gives rise to one of the most approachable demonstrations of Pythagoras' theorem. As a is the side of the square (the hypotenuse of each sponge) and b and c the perpendicular sides of the sponge (the other two sides of the triangle), we find that the area of the square is broken down into:

$$a^2 = 4 \cdot \frac{b \cdot c}{2} + (b-c)^2.$$
$$a^2 = 2bc + b^2 - 2bc + c^2.$$
$$a^2 = b^2 + c^2.$$

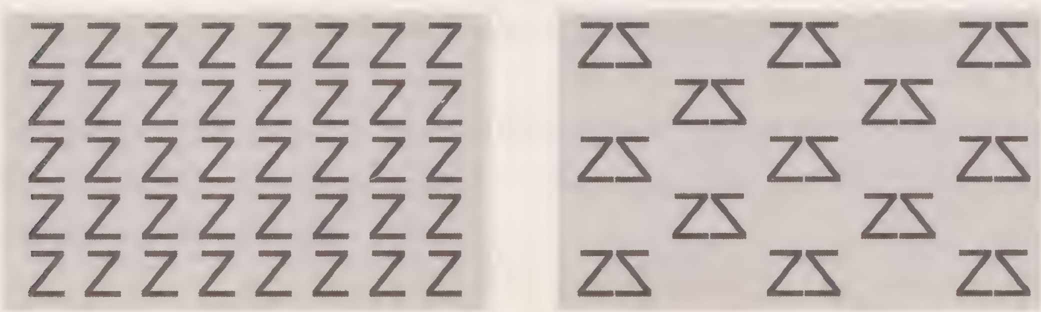
Variations on themes

Musical composers know that the repetition of a motif creates unity, but the way in which it is created can generate monotony and boredom. The joy in a good composition, be it musical or of any other type, can be based more on the creativity shown than on the repetition of the motif itself.

Designers are also faithful to this idea when the logo for a brand is repeated indefinitely to decorate their products and bottles. The most commonly-used resource is symmetry. Symmetry does not alter the shape, it only changes the place and position, as there are three transformations on a plane which conserve shape: translation, rotation and reflection.

Translation changes the position. Rotation rotates a figure around a point, the centre of rotation. The reflection is an axial symmetry, in other words, a reflection in a mirror. With these three transformations, one motif can be repeated, creating up to 17 patterns which are structurally different from a mathematical point of view. Creative design does not require this many.

For example, for the brand *Z*, we could design repetition patterns like those shown below. It can be seen that in the repetition the letter loses its literal meaning in order to adopt the figurative character imposed on it.



The advantage of this system resides in the easy identification of the brand and its ease of systematic reproduction. There are practically no brands that do not use ornamental patterns on their products.

Epilogue

Manual for Mathematical Creators

Is it possible that someone could think that Stravinski's *The Rite of Spring* or *Guernica* by Picasso were discoveries? What is it that causes mathematical discoveries, in most cases, to be taken as discoveries and not creations? Is Pythagoras not attributed with the creation of the diatonic musical scale? Why is he seen as the creator of that but not of the proof of the theorem that carries his name? Neither the diatonic scale nor the demonstration of the theorem were hidden among the foliage or inside a cave waiting for a discoverer; they were both creations of the same man. And, in case you are in any doubt, try to discover the celebrated theorem offered in Euclid's *Elements* for yourself.

The British mathematician Andrew Wiles gives a great description of what a mathematician feels from the moment of encountering a problem that he cannot solve to finding the solution:

> "Perhaps I could best describe my experience of mathematics in terms of entering a dark mansion. One enters the first room of a mansion and it's dark. Completely dark. One stumbles around bumping into the furniture, but gradually you learn where each piece of furniture is. Finally, after six months or so, you find the light switch, you turn it on, and suddenly it's all illuminated. You can see exactly where you are!"

Not everyone in the world has the intellectual capacity of Wiles, who was able to prove a theorem, one of Fermat's, which had eluded mathematicians for more than three centuries. Perhaps the greatest mathematicians have some common strategy to resolve problems and demonstrate theorems.

The first lesson consists of daring to enter the dark mansion. Many people underestimate their ability to tackle a problem. You have to dare and be prepared to bump into furniture. Contrary to appearances, and this is something that mathematicians and educators have consciously hidden – mistakes help. By bumping into furniture you learn where things are. And even though you may not see the

room in full light, you end up with an idea of what it is like and its arrangement. It is then that you know you are ready because you know what you are up against. You have reached the point where you can clearly formulate the problem. You are still moving around in the dark, but your thoughts have been illuminated.

A recap

What are the characteristics of mathematical creativity? Creating mathematics is, above all, having good ideas for finding routes to new formulae, theorems and procedures with which, little by little, we gain an understanding of the phenomena we experience and which concern us.

Archimedes' Eureka!, the sudden inspiration explained by Poincaré which many call the 'happy idea', often forms part of a creation. But it is rarely the fruit of chance or luck. Instead it appears after a period of intense continuous work (the period of incubation) which psychologists class as a phase in the creative process known as 'illumination'. Heuristics, the art of invention and discovery postulated by George Pólya and Imre Lakatos, provides guidance to being creative in mathematics. Logic itself is not creative, but it is indispensable, above all in the essential phase of verifying the hypothesis.

Crisis and change go hand in hand with great mathematical creations. Crises are overcome with change. The development of new knowledge means change that is more difficult to accept than it seems, since new concepts are often, in a certain sense, 'monsters' which imply a profound revision of previous knowledge. This was the case with powers with a negative exponent, the square root of 2 and the square roots of negative numbers. Euclidean geometry also lost its exclusiveness with the creation of other geometries. The most recent of them is fractal geometry, developed as recently as the second half of the last century thanks to technological advances. Until then the mathematician's luggage had been very light: paper and a pencil were all that was needed. Nowadays, the computer has become essential in many fields of mathematics.

One way of introducing newcomers to mathematical creation is the use of tools which are inherent to the discipline to explore different phenomena, both in the personal sphere and in art, literature and work. Looking in the mirror or contemplating the horizon led to the creation of formulae and relationships with which we can learn

how to improve those situations. None of those formulae were discovered, they were not found behind glass or buried in the sand on the beach. Rather, they were the product of our imagination to model reality with geometric objects. Without us there would be no formulae and no horizon. Among the cases we looked at were situations taken from artistic and working practices. In the first case, craftsmen and artists have often found themselves needing to work with sufficient geometric accuracy so that their work has the desired appearance. This explains the fact that interwoven knots, which are so common in cultural decorations, are created on grids. The enigma of the cyclic and infinite knots of Celtic folklore resides in a matter of divisibility. The theorem did not create the knot; the knot created the theorem.

In the world of work, there is an abundance of applications of mathematics in the resolution of a large number of issues. Once again we find situations in which the theoretical solution to a practical problem may not be the best practical solution to the problem. By studying diverse problems we have seen that there is no one way to conceive an equilateral triangle, that the sum of the rounded parts is not the rounded sum and that, except in extraordinary cases, a third of what we see is not a third of what we are looking at. We can also find centres of creation and mathematical inspiration outside of mathematics. Argentinian writer Jorge Luis Borges created two fictional devices that both constitute strange and impossible, but fascinating, mathematical objects: a book with infinite pages and a disk with just one face. We proposed a geometric problem and a new theorem based on a sentence written by Italo Calvino.

Addressing Varignon's theorem it becomes patent that a logical demonstration cannot explain the phenomenon, and a look at the context is needed to understand it. This strange example of the importance of the 'environment' of any problem provides us with a wonderful metaphor for addressing another of the book's fundamental theories: in short, mathematics is created not just by professional mathematicians, not only in universities, not only in research centres and not only in our Western culture. The story of how the Indonesian craftsmen learnt to design regular polygons through non-Euclidean methods demonstrates the relevance of the social and cultural aspects in mathematical creation. Experiencing unusual things opens the mind and makes us creative because we have to decipher new ideas with what we already know. Among them is that of interpreting a bamboo stick as a piece of software.

Advertising and design have become test beds for innovation. When strategies have run out and when it seems that things have been approached in every possible way, new ideas appear that take their inspiration from old, but less common ideas. Mathematics sells. Anyone who captures or understands the message in an advert or appreciates the beauty and perfection of a geometric design is not stupid. Mathematical objects and concepts have come to form part of our everyday life thanks to their use in advertising campaigns and the design of objects that are available to everyone. Numerical and figurative proportionality, Möbius strips, the Pythagoras' theorem, symmetry, infinity, functions and their graphs, probability, etc. Which elements of mathematics are yet to be used in design and advertising?

Throughout history two fundamental types of creation can be seen which, as always tends to happen, are interdependent: the creations of expansion and of assimilation. Expansion means extending known terrain, moving towards the horizon. With the term assimilation, we are referring to creations that have meant a change in conception of the fundamentals of mathematics. Creations of this type are those that have appeared in the great crises.

Guidelines for mathematical creation

Creation consists of making something exist which was not there before, be it an object, a procedure or a concept. Creating mathematics basically consists of resolving problems. Therefore, the first step in creating is proposing a question which mathematics can answer. Anything can be used as inspiration, whether it is a theorem, another mathematical problem or an everyday activity. Once immersed in the matter, it is necessary to look closely to see common, underlying patters in similarities or differences. Social and cultural interaction can be very relevant in this process. Sharing some of our ideas and failed attempts can help us to illuminate the dark room in which we stand. Technology can also be a great help and inspiration, as it makes it possible to visualise nearly everything imaginable.

Finally, we will never tire of insisting that the object of mathematical creation consists fundamentally in offering explanations for understanding a phenomenon. Logic validates, but it does not always explain. Fermat's theorem was demonstrated in the end, but the understanding of the phenomenon which it proposes relies on such varied arguments and analogies that it does not explain its essence. That is why we still miss Fermat's supposed marvellous proof. And we do so because we continue

to think that there could be a route that is not only shorter, but clearer. We are keen to comprehend, and we only long for the demonstration having understood the phenomenon.

Throughout this book the idea that everyone can create mathematics has been upheld, that mathematics, in short, is a much more democratic activity than history (or certain parts of it) and educational tradition would have us believe. So what is the key, the method, that can make us into creators, creators of mathematics? From everything that has been demonstrated we can draw a few fundamental conclusions:

1. Propose mathematics questions through life experiences.
2. Carry out a mathematical study of the matter.
3. Propose the question in mathematical, or quantifiable terms.
4. Use resources such as experimentation, intuition, analogy, logic, as well as technological and cultural tools.
5. In other words, live mathematically, at least for part of every day.

Happy creation!

Appendix

Parallel Lines that Intersect on the Plane

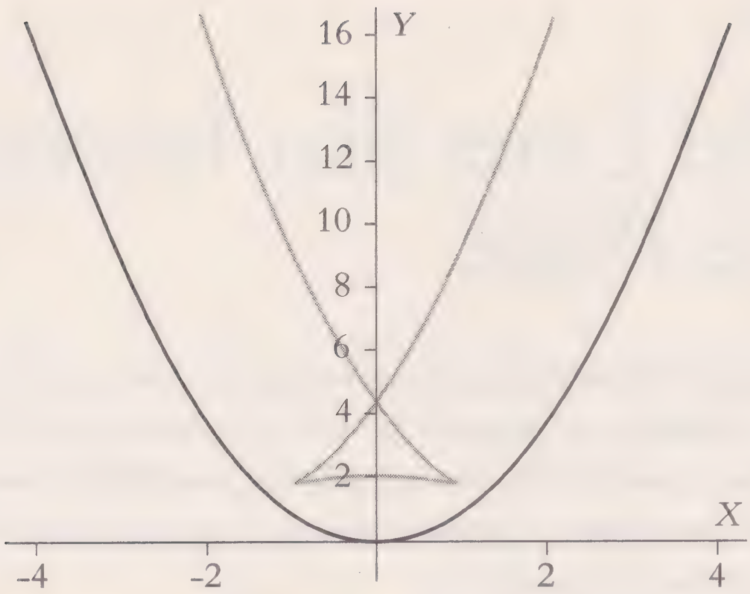
This is an example of how the use of technology can upset deeply rooted preconceptions. It is included in an appendix in order to be able to include the somewhat complex symbolic notation. The idea of parallelism is unfailingly and unconsciously associated with the concept of parallel segments or straight lines. We tend to think of parallel as automatically rectilinear.

A while ago I had an experience that led me to look for the expression for the curve parallel to a given curve $C(t)=(t,f(t))$ at a distance d. Why? Because the bus on which I was travelling had to stop because a tree had fallen across the road. It's trunk was so big that it was impossible for the bus to pass over the top of it. When imagining the situation as seen from above I saw two parallel lines (the road's drainage ditches) with a segment perpendicular to both of them (the fallen tree). It was then that I asked myself what the expressions for two parallel curves on the plane would be. The result was the following parametric equations:

$$x(t)=t-\frac{d\cdot f'(\mathrm{t})}{\sqrt{1+f'(t)^2}}.$$

$$y(t)=f(t)+\frac{d}{\sqrt{1+f'(t)^2}}.$$

At this point I wanted to check experimentally if those formulae worked as they were supposed to. To do so I used a computer program with which I wanted to see the original curve along with various other parallel ones at different distances. I started with the simplest of the curves, the quadratic curve $y=x^2$, that is, curve $C(t)=(t,t^2)$. The result was as expected for distances such as $d=1$, $d=2.5$, $d=0$, $d=-0.5$. But when increasing the negative values of distance d I came across the following:



What was happening? I must have made some sort of error, as the parallelism contradicts the intersection. Parallel lines do not intersect! I went over the calculations and saw that the formulae were correct, as were the instructions I had written so that the program carried out the graphical presentation correctly. Surprise! On my computer screen there were parallel lines that intersected.

Definition no. 23 of 'Book 1' of *Elements* states: "Parallel straight lines are straight lines that, being in the same plane and being produced indefinitely in both directions, do not meet one another in either direction". The idea of parallelism is so contextualised on the plane that we take the impossibility of intersection as a paradigm of parallelism. My curves had been drawn on a plane, but I had not conceived them according to that reference point, but in terms of equidistance. For straight lines on the plane, both things are the same, but not for curves. Parallel curves can cross.

Although, thinking about it, do the Euclidean definitions and postulates exclude the possibility of considering circular arcs as straight lines? If not, two circular arcs with different radii and the same centre could be extended in both directions without ever intersecting and we would have a pair of parallel curves. And, what is a straight line for Euclid? A straight line is "a line that lies evenly with the points on itself". Does a circular arc not lie in the same way in terms of it points? The Euclidean definition of the straight line was undoubtedly not created with circular arcs in mind, but circles do fit with it. Another option would be to take Archimedes

definition as a reference (although it generates other straight lines): "The straight line is the shortest of all the lines with the same two ends."

In any case, nothing stops parallel curves from intersecting. My prejudices were derived from the parallelism of straight lines. In the case of straight lines on the plane, the notion of parallelism can be substituted by that of non-intersecting straight lines, but not in the case of curves. Parallel curves can cross and, in fact, they do.

I later created a dynamic interpretation of this phenomenon: I imagined the distance as a rigid segment which, when reaching the vertex of the parabola, has to manoeuvre like a car to describe a curve that is too closed. That is, it stops and goes into reverse, if necessary, without modifying the width of the road, which is the width of the parallelism.

A while afterwards I created a three-dimensional version of the problem of the parallelism of curves. If two parallel lines exists on a given curve for each predetermined distance, in space they are infinite in number, and form a tubular shape with the original curve in its circular centre.

Bibliography

BOYER, C.B., *A History of Mathematics,* Hoboken, John Wiley, 1991.

COURANT, R. and ROBBINS, H., *What is Mathematics? An Elementary Approach to Ideas and Methods*, edited by Ian Stewart, Oxford University Press, 1996.

DAVIS, P. and HERSH, R., *The Mathematical Experience*, New York, Houghton Mifflin Harcourt, 1999.

ERNEST, J., *Social Constructivism as a Philosophy of Mathematics*, Albany, Suny Press, 1991.

HERSH, R., *What is Mathematics, Really?*, New York, Oxford University Press, 1997.

LAKATOS, I., *Proofs and Refutations. The Logic of Mathematical Discovery,* Cambridge, Cambridge University Press, 1976.

POINCARÉ, H., *Mathematics in the Modern World,* Scientific American, 1974.

PÓLYA, G., *How to Solve It: A New Aspect of Mathematical Method*, Princeton University Press, 1988.

Index

Creative Mathematics

The amazing workings of the mind

Anyone and everyone can create mathematics! Using historical and contemporary examples from societies around the globe, this book demonstrates that mathematics is a democratic activity. The reader will discover the endlessly beautiful simplicity and complexity of mathematics, and learn that the secret to a lively mind is to "live mathematically" a little every day.